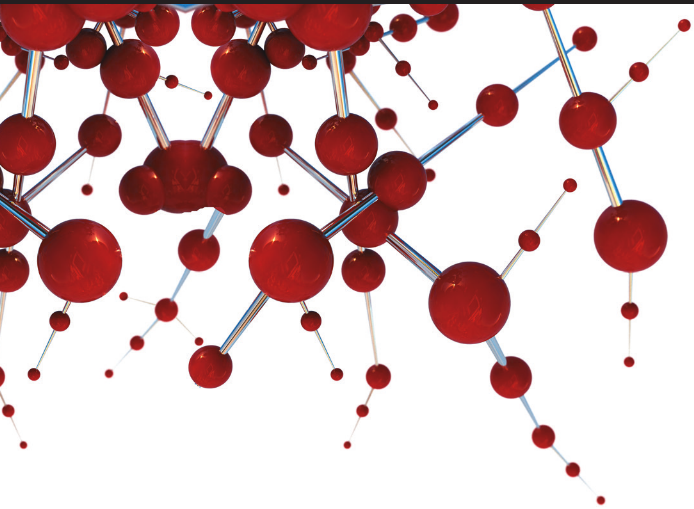
Interactive Technology · Faculty of Information Technology and Communication Sciences · Tampere University



Csaba Kertész

Collaborative Artificial Intelligence Development for Social Robots

Csaba Kertész

Collaborative Artificial Intelligence Development for Social Robots

ACADEMIC DISSERTATION
To be presented with the permission of the Faculty of Information Technology and Communication Sciences of Tampere University, for public discussion in the auditorium Virta 109
on May 27th, 2022, at noon.

Faculty of Information Technology and Communication Sciences
Tampere University

Dissertations in Interactive Technology, Number 36
Tampere 2022

Academic Dissertation in Interactive Technology

Supervisor:	Professor Markku Turunen, Ph.D. Faculty of Information Technology and Communication Sciences, Tampere University, Finland
Opponent:	Mohammad Obaid, Associate professor Department of Computer Science and Engineering Chalmers University of Technology Gothenburg, Sweden
Reviewers:	Dr. Amit Kumar Pandey, Chief Scientist/CTO beingAI Hong Kong, China Christoph Lutz, Associate Professor Department of Communication and Culture BI Norwegian Business School Oslo, Norway

The originality of this thesis has been checked using the Turnitin OriginalityCheck service in accordance with the quality management system of Tampere University.

Dissertations in Interactive Technology, Number 36

Faculty of Information Technology and Communication Sciences
FIN-33014 Tampere University
FINLAND

ISBN 978-952-03-2442-1
ISSN 1795-9489

Juvenes Print – Suomen Yliopistopaino Oy
Tampere 2022

Abstract

The main aim of this doctoral thesis was to investigate on how to involve a community for collaborative artificial intelligence (AI) development of a social robot. The work was initiated by the author's personal interest in developing the Sony AIBO robots that have been unavailable on the retail markets, however, user communities with special interests in these robots remained on the internet.

At first, to attract people's attention, the author developed three specific features for the robot. These consisted of teaching the robot 1) sound event recognition in order to react to environmental audio stimuli, 2) a method to detect the underlying surface under the robot, and 3) of how to recognize its own body states. As this AI development proved to be very challenging, the author decided to start a community project for artificial intelligence development. Community involvement has a long history in open-source software projects and some robotics companies tried to benefit from their userbase in product development.

An active online community of Sony AIBO owners was approached to investigate factors to engage its members in the creative processes. For this purpose, 78 Sony AIBO owners were recruited online to fill a questionnaire and their data were analyzed with respect to age, gender, culture, length of ownership, user contribution, and model preference. The results revealed the motives to own these robots for many years and how these heavy users perceived their social robots after a long period in the robot acceptance phase. For example, female participants tended to have more emotional relation to their robots than male who had more technically oriented long-term engagement motivation. The user expectations were also explored by analyzing the answers to this questionnaire to discover the key needs of this user group. The results revealed that the most-wanted skills were the interaction with humans and the autonomous operation. The integration with the AI agents and Internet services was important, but the long-term memory and learning capabilities were not so relevant for the participants. The diverse preferences for robot skills led to creating a prioritized recommendation list to complement the design guidelines for social robots in the literature.

In sum, the findings of this thesis showed that developing AI features for an outdated robot is possible but takes a lot of time and shared community efforts. To involve a specific community, one needs first to build up trust by working with and for the community. Also, the trust for the long-term endurance of the development project was found as a precondition for the

community commitment. The discoveries of this thesis can be applied to similar types of collaborative AI developments in the future.

There are significant contributions in this dissertation to robotics. First, the long-term robot usage was not studied on a years-long scale before and the most extended human-robot interactions analyzed test subjects for only a few months. A questionnaire investigated the robot owners with 1-10+ years-long ownership in this work and their attitude towards robot acceptance. The survey results helped to understand the viable strategies to engage users for a long time. Second, innovative ways were explored to involve online communities in robotics development. The past approaches introduced the community ideas and opinions into product design and innovation iterations. The community in this dissertation tested the developed AI engine, provided inputs for further development directions, created content for the actual AI and gave their feedback about product quality. These contributions advance the social robotics field.

Acknowledgements

First of all, a special thanks to my supervisor, Markku Turunen, who supported me in pursuing my doctoral degree over the years. I would like to thank all fellow students and researchers at Tampere University for the creative discussions and the feedback for my ongoing work. I appreciate all lecturers' efforts at my university and online course platforms to complete the planned courses during my studies. Thanks to all researchers worldwide who did not hesitate to share their research data when I asked to help the common scientific advance. Thanks to the Nokia Foundation for the research grant in 2015 to boost my initial efforts in this journey. Thanks to my family members for being patient when I worked on my research. And the whole story did not happen if my mother, my deceased father and my sisters would not support from the beginning since I am on the Earth. My grandpa gave the inspiration for my Ph.D. studies. Let it be the will of God.

Tampere, April 12, 2022

Csaba Kertész

Contents

List of Publications

This dissertation is composed of a summary and the following original publications, reproduced here by permission.

I. Kertész, C., & Turunen, M. (2018). Exploratory Analysis of Sony AIBO Users. AI & Society, 1−14. doi:10.1007/s00146-018-0818-8

II. Kertész, C., & Turunen, M. (2017). What Can We Learn from the Long-Term Users of a Social Robot? In Proceedings of the 9th International Conference on Social Robotics, 657−665. doi:10.1007/978-3-319-70022-9_65

III. Kertész, C., & Turunen, M. (2018). Common Sounds in Bedrooms (CSIBE) Corpora for Sound Event Recognition of Domestic Robots. Intelligent Service Robotics, 11(4)335–346. doi:10.1007/s11370-018-0258-9

IV. Kertész, C., (2016). Rigidity-Based Surface Recognition for a Domestic Legged Robot. IEEE Robotics and Automation Letters Journal, 1(1)309-315. doi:10.1109/LRA.2016.2519949

V. Kertész, C., & Turunen, M. (2018). Body State Recognition for a Quadruped Mobile Robot. Proceedings of the IEEE 22nd International Conference on Intelligent Engineering Systems, 323-328.

The Author's Contribution to the Publications

The publications in this thesis were coauthored, the author was the main contributor, responsible for planning, executing the experiments, data analysis and writing the papers. The coauthor was in a supporting role, involved in commenting the papers, advising the research directions and guiding the work. Publications I and II would not be possible if the enthusiastic robotics community members do not fill the analyzed survey without any compensation.

1 Introduction

Artificial intelligence and deep learning are recent innovations at the forefront of the robotics industry, changing the world, society and the job market. The accelerated development of these technologies made the robots more advanced and the neural networks surpassed the human capabilities in some domains (e.g. playing games, recognizing faces). The social robots are such autonomous robots which are geared towards human-robot interactions. These robots are physically embodied with a strong emphasis on their social behaviors during their interactions. This thesis focuses on the collaborative development of social robots.

Despite the fact that intelligence has a varying definition in the research fields (see Section 1.6), artificial intelligence (AI) aims to replicate human intelligence by a machine and it has two main branches. The artificial general intelligence (strong AI) builds general-purpose machines while the narrow artificial intelligence creates machines for a specific task. Regardless of the problem, an AI solution involves reasoning, mimicking human thought processes and problem-solving. One aspect of social robots is to implement a narrow intelligence to induce an emotional attachment with their owners and improve their mood by interactions. These robots exhibit thinking and rational behaviors according to the expectations of the typical human-robot interactions.

Nowadays, the narrow AIs solve particular problems and these AIs get the fame in the media. Notable examples are Amazon Echo, Siri and iRobot as materialized products based on advanced AI technologies. However, artificial intelligence development is challenging. The latest technologies are not mature enough to develop robots or virtual agents, giving an impression of human-level intelligence. Some tricks can lower the expectations of humans (e.g. resembling an animal), but the academic researchers and robotics startups face an overwhelming problem. Building a strong AI is not viable today though a general-purpose AI can be approximated by fusing multiple AI methods into one AI engine. The difficulty in this process is the increasing complexity after including more and more narrow AIs.

Since artificial intelligence development is a challenging task even with the latest technologies, a natural solution can be online communities' involvement for the common goal of a successful social robot. To crack this problem, the existing examples inside the robotics industry are examined in this dissertation and an open-source project run by the author is described to engage community members. The author believes that collaborative development with online communities is one possible fix because external human resources can reduce the burden of the AI engine's exploding complexity.

1.1 GOAL

This thesis investigates how to involve online communities in the artificial development of social robots because the existing literature on this topic is scarce. Some works (McAlexander et al., 2002) (da Mota Pedrosa, 2012) examined collaborative open-source software development and online communities' involvement in innovation and product designs. However, there have not been similar studies in the robotics and artificial intelligence domains. This research aims to fill this gap by addressing the following research questions:

RQ1: How long-term owners of a social robot can be involved in collaborative artificial intelligence development?

RQ2: How can the future social robotics projects run collaborative artificial intelligence development successfully?

The author joined an online community of Sony AIBO robot owners. He participated in the everyday discussions and executed a questionnaire among the members to get know the target audience to answer RQ1. This survey reached the active members on the forum, it assessed the demographics, their skills for AI development and their expectations for new software. This data revealed how these members could be engaged for contributions and their skills were found out to assign appropriate tasks for them. Two publications described the analysis of this questionnaire and general recommendations for social robot design based on the findings. These analyses were important to develop a basic AI engine to get the community members' attention for involvement and contribution. The other three publications in this dissertation describe the technical background of these AI features. After this project was run for some years, important empirical observations were gained and they are discussed in the context of the past examples in the robotics industry. The author believes that putting these results in a bigger picture is beneficial for executing future social robotics projects by maximizing collaborative development potentials.

The answers to the other research question (RQ2) can be deducted from the outcomes to RQ1. The reasons for failures in the social robotics companies are discussed and compared to the experiences of this thesis work. The design recommendations in Publications I and II are direct help for future projects to design successful robotics products. In simple terms, a new robot must follow the technological constraints and minimize the expectations to a reasonable level. The company should not set unrealistic goals and close communication to the early adopters reveals challenges enough early before entering the mass market. If the company aims to release a product with advanced intelligence, the roadmap must contain resources for the continuous content creation to keep the consumers satisfied. The collaborative development can provide a partial solution to outsource some efforts to the community, but a stable development environment with good documentation and an online forum is a must for such plans.

1.2 METHODOLOGY

To the author's best knowledge, there was no similar research in the literature before this work was executed and five articles were published on this topic. Each publication was related to a scientific subfield and followed the scientific methodologies of those subfields. These articles were part of a case study in which the author executed a collaborative AI development with an online community. This project provided some valuable insights to run a real-world experiment in the wild instead of a laboratory setting. The gained experiences were set against the social robot industry examples as a comparative study to enhance future social robot projects.

This dissertation is a hybrid work regarding scientific methodologies reviewed in (March & Smith, 1995). On one hand, the community of robot owners was queried by a questionnaire for analysis. This method originates in the traditional social science that is listed like a *natural science* in (March & Smith, 1995). The primary goal was to get know this user population in reality. This approach helps to identify the target group of the social robotics industry since the real customers in our society were never examined by academic research in this depth before. This part of the dissertation work is exploratory, it also studies if the existing theories in the literature are valid for the long-term Sony AIBO users and it proposes recommendations for social robotics research and industry. On the other hand, the dissertation is practical when the author tries to engage the community members for contributions to artificial intelligence development. The author implemented several AI features to the robot to get the attention of the community members and several articles resulted by this *design science* work. Those AI features had the direct utility to serve the author's purposes to improve the chances for collaborative development with other community members. In the next paragraphs, the scientific methodology of Publications I-II describes the natural science part of the dissertation and the Publications III-V are shown as examples for design science.

Publications

Publications I-II used social science methods to analyze robot dog owners of an online community. Since these robots were not sold for years on the retail market, these people were long-term robot owners. These community members filled out a questionnaire to get know why they kept their robots for so long and what are their expectations for a new AI for their robots. Their responses were quantitative and free-form text. This survey had many questions and different methods were used for the quantitative and free-text answers, the former was investigated in Publication I and the latter in Publication II. Common statistical methods evaluated the answers to the Likert-type items and the text answers were grouped together by topics to draw conclusions from their prevalence. After the analysis was finished, both articles formulated recommendations for social robot design.

Publications III-V were technical articles about the implementations of some AI skills to attract community members' attention. Publication III experimented with sound event recognition for the robot to react to environmental audio stimuli. Publication IV described a method to detect the underlying surface under the robot and Publication V taught for the robot how to recognize its own body states. These three articles shared a common methodology of machine learning evaluation. First, a dataset was collected with a robot to represent the desired task for each article. The relevant features were extracted from the dataset for machine learning training, the dataset was split into training and validation sets. Several classifiers were evaluated by cross-validation on the training set and the validation set was shown to them as unseen data to find the best performing. Once the best classifiers were identified, the real-time capabilities of every classifier were taken into account for deployment on the robot. After the trade-off between accuracy and real-time performance was examined, a final classifier was selected and the model was incorporated as a new skill into the AI engine.

AiBO+ Project

This paragraph describes the executed community project in short. After the author realized the challenges in artificial intelligence development, he explored solutions to these problems. He decided to start a community project for artificial intelligence development, but he knew that the hardware and software developments are hard on their own. Since he was more passionate about writing programs, he looked at the ready-made robotics products with an open software development environment and low-level access to control the hardware. Sony AIBO robot dogs were chosen because they satisfied these criteria, they were affordable compared to other complex robots and they could move around while having a lot of sensors onboard. The best model of this brand was the Sony ERS-7. The author bought this model on the secondhand market in 2008 and started to code an initial AI engine for the robot for some years to use it for the engagement of community members. The project was named AiBO+, the first public software was released in 2015 and active communication was established to a community of Sony AIBO owners on an online forum. The dissertation covers these collaborative efforts during 2014-2020.

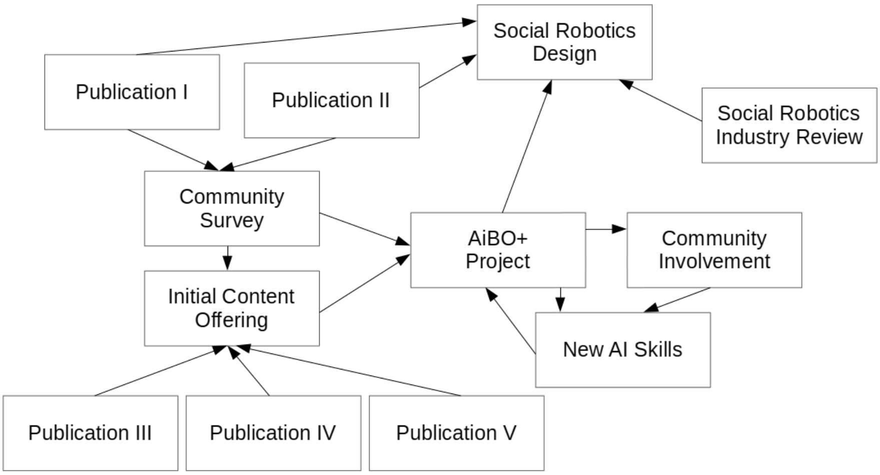


Figure 1. Dissertation overview. This diagram shows how the publication results supported the collaborative development and concluded in social robotics design recommendations.

Summary

Figure 1 shows the overall structure of the efforts in this dissertation. On the one hand, Publication I and II provided inputs for the AiBO+ project via a community survey to identify the target community and the wished AI skills for a successful engagement. On the other hand, the survey results in those publications suggested good practices for designing social robots which are general outcomes of this dissertation. After the necessary actions were pinpointed for user engagement, these findings were implemented in software for the robots to get the attention of the community members and the project was announced publicly. Publication III, IV and V detail the technical concepts for this initial content offering, scientifically evaluated artificial intelligence features that work in practice. The ongoing AiBO+ project engaged several participants and the AI development was executed in a continuous feedback loop, similar to the standard agile software development model. The community survey gave a general basis to the future planning in a middle run, the active contributions were included in the upcoming AI engine updates while the community feedback was taken into account retrospectively. Finally, the community project experiences, a review of the social robotics industry and Publication I/II accumulated design recommendations for forthcoming social robotics projects in academia and the industry.

1.3 RESULTS

The thesis proves that even old robotics hardware can be utilized for cutting edge research. The author could implement a basic artificial intelligence by modern software and industrial software development practices. Since the available technologies are limited to create intelligent robots, there are many open research questions that can be explored with constrained resources and budget. Another result was the engagement of multiple community members to contribute to the AI development despite the AiBO+ project was driven by the author alone.

The targeted community was discovered by a questionnaire during the experiments and the developed AI features were rigorously verified by scientific methods. These contributions to the overall results were published in five articles. Publications I-II showed an analysis of long-term social owners. This part of the research is unique because such robot owners were not examined before who actively use their robots for years (heavy users). While the primary goal of these articles was to get know the target audience, the analysis of this community suggested recommendations for social robot design. Publications III-V reported the implementation of several AI skills to engage the community members for contribution. On one hand, these technical solutions were explored and verified by rigorous scientific methods in the articles. On the other hand, the developed skills were included in the AI engine and they were distributed among the aforementioned robot owners freely.

1.4 CONCLUSIONS

This research was novel since the past works on collaborative development examined open-source projects and product innovations, but this dissertation looked into the artificial intelligence for social robots. An important conclusion was the essential influence of the robot design for collaborative work and consumer expectations. Robotics projects must build their roadmap upon careful considerations regarding the aesthetics and the companies need to envision the subscription-based products to found the continuous software development for their customers. Most robot owners are non-technical people thus the collaborative development plans have to consider their possible contributions. The execution of AiBO+ project was successful because code, content and scientific contributions were received for the AI development without remunerations.

1.5 Future Work

This dissertation was an exploratory work to involve online communities to artificial intelligence development. The executed survey provided sufficient information for the AiBO+ project to schedule the AI skill development and get know the community members. Nevertheless, the literature lacks similar long-term follow-up studies to track robot owners. The university labs should build connections with social robotics companies to query their customers frequently how their attitude changes towards their robot over time. Other future studies can be designed to replicate the experiences in this research or analyze the implementations of the social robot recommendations in future robotics projects.

1.6 On Terminology

Online community

The online community is a virtual community on the internet where physically distant people (members) interact with each other. The members participate in these communities because of a shared goal. Their reasons vary like posting, commenting on discussions, giving advice or collaborating for a common target. The online communities are used in two primary ways in this dissertation. On one hand, the author refers to Sony AIBO owners on an online forum, on the other hand, the phrase can mean software developer communities around open-source projects.

Artificial Intelligence

The definition of artificial intelligence originates in the exhibited intelligence found in nature (human, animal) and the machines mimic this natural intelligence. The artificial intelligence is usually used in a broad sense, describing any human-written program or algorithm which can be considered intelligent. Unfortunately, there is no established definition of intelligence in robotics or psychology. Therefore, the author mentions artificial intelligence as an umbrella to describe any program or machine learning method which has some learned intelligence or problem-solving capability.

Social robots

Social robots are robots with a specific purpose although they are powered by the same artificial intelligence technologies. The primary context for social robots is the human-robot interaction. They need to understand humans, carry out a conversation intelligently and perform various tasks if they are asked for. For example, the robot used in this dissertation is a Sony ERS-7. This social robot does not have a direct utility, but it was designed to entertain people and act as a companion.

Open-source software

The open-source software is a computer program whose source codes are accessible by 3rd party developers. It is released under a free license to grant the rights to download, use, change and redistribute the compiled software with no or minimal restrictions. Normally, open-source software is hosted on a website with version control and other software coder services. The free software receives contributions from 3rd party developers and it is developed in a collaborative public manner. There are more definitions of open-source software depending on the openness level of the attached license. The author uses this term for any software which is available on the Internet for collaborative development and it has a free license to enable 3rd party contributions.

Collaboration

Collaboration involves two or more people working together to finish a task or accomplish a goal. As the definition of open-source software and online community already showed, there is a relationship between them and collaboration in the context of this dissertation. Collaboration is used in this thesis as a cooperative action between online communities' members to develop an artificial intelligence engine, regardless of the contribution type (technical or non-technical).

Heavy user, long-term owner

The heavy user and long-term owner expressions are used in an interchangeable way in this thesis. They describe such customers who purchase and own their robots for a longer time period. There is no exact definition in the literature on what long-term usage means in robotics research. The author considers robot owners as heavy users after one year of ownership because the initial wow-effect already faded away for these users and they do not abandon their robots because of emotional bonds.

Robot acceptance

The technology acceptance is a process for a consumer to acquire and use a product, nonetheless, the same process is applicable for robots (M. M. A. de Graaf et al., 2014). When a person gets know about a robot, it is the pre-adoption phase. After the purchase, the customer will have the initial wow effect in the adoption phase. After 1-2 months, the person gains deeper experiences with the robot in the adaptation phase and the robot will be part of the daily routines in the incorporation phase. And finally, the robot will be accepted as a personal object for the consumer after some more months in the identification phase. As the participants of this dissertation project were heavy users and they owned their robot for more than a year, they were already in the robot acceptance phase.

Skit

The skit is a special term in the terminology for Sony AIBO. It means a collection of motor motions, LED animations and sound played back at the same time. A typical skit is a dance where the robot plays a music hit, dances and flashes the LEDs on its head to the rhythm.

2 Collaborative Development and Social Robotics

The subject of this dissertation is the collaborative AI development for social robots hence the available literature is reviewed in this chapter. First, the robots are introduced and the social robots are described how they are a unique genre inside robotics. The typical usage of social robots is covered next, along with the target audiences of these robots and the requirements in the human-robot interactions. After some example social robots are presented from the mass market, the current status and challenges are reviewed in the social robotics industry. The second half of this chapter deals with collaborative development that has a long tradition in software development. Although online communities can collaborate in different ways, their role in product design and innovation is a common research direction. The open-source development's major pillars are looked through and how these software development processes can be engaged inside an online community. Furthermore, a robotics company is analyzed how it utilizes collaborative development in business operations.

2.1 SOCIAL ROBOTS

The term “robot” was invented by Karel Capek, a Czech writer in 1921. The earliest industrial robots were designed in the 1930s and the humanoid robots appeared at universities in the 1970s. After the rapid development in computer sciences since the 1980s, the robot capabilities were improved significantly, their costs decreased and they entered into the commercial market (Breazeal, 2017). The first successful robots were developed to implement one function, for example, iRobot released a vacuum cleaner robot Roomba (Jones, 2006) and Sony made a line of entertainment robots

(Sony AIBO). The increasing number of robots among us projects that robots in our society can be ubiquitous after a few decades, similar to computers and smartphones today (Šabanović, 2010). Since this prediction is likely to come true, the robots' scientific, technological and economic impacts must be studied (Innes & Morrison, 2017).

A scientific definition can describe the robots as multifunctional, programmable tools that can manipulate objects and respond to the changes in their environment in order to perform tasks (Hockstein et al., 2007). Most robots replace humans in dangerous situations (Zhuang et al., 2008) or do repetitive (Prassler et al., 2000), precise (Hempel et al., 2003), tedious tasks (Cormier et al., 2013) and other robots are targeted for mental support (Riek, 2016), entertainment (Fujita et al., 2000) or being a companion for humans (Vu et al., 2015). These latter robots are called social robots because their operation is based on the same robotics principles although in a social interaction context (Salichs et al., 2006). They execute simple, supportive tasks in various environments like helping a factory worker by tool handling in a cooperative fashion (Weiss et al., 2011) or serving the elderly in nursing homes (Rantanen et al., 2017). Thus, social robots work together with humans in homes (Wilson et al., 2019), hospitals (Takahashi et al., 2010) and educational institutions (Belpaeme et al., 2018). People expect communicative skills from intelligent robots in general (Burget et al., 2017), but cooperation does not require extensive dialogs necessarily (R. Liu & Zhang, 2019). And it is not directly intuitive though, but uncooperative behaviors can make social impressions like playing a game with humans competitively (Horstmann & Krämer, 2020).

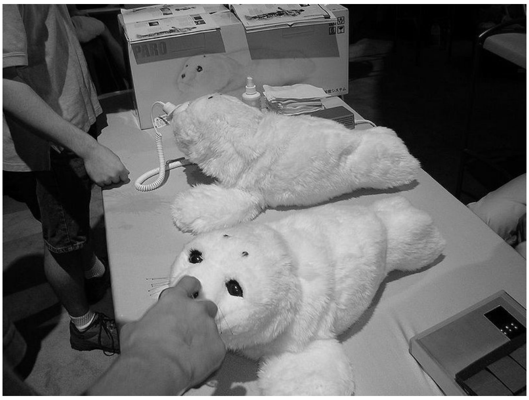


Figure 2. Paro robot. (License: Creative Commons Attribution-Share Alike 2.0 Generic. author: Aaron Biggs, source: https://commons.wikimedia.org/wiki/File:Paro_robot.jpg)

Several social robots have been available commercially in the past two decades, for example, Paro (Šabanović et al., 2013), Pleo (Fernaeus et al., 2010) and Sony AIBO (Fujita, 2000). Unlike the other robots on this list, Paro (Figure 2) was intended for healthcare use (Hung et al., 2019). This robot resembled a baby seal and it responded to petting and cuddling to result in a calming effect and responded to the users emotionally (Geva et al., 2020). The target groups of this robot were older people in nursing homes and patients with severe illnesses in hospitals (Šabanović et al., 2013).

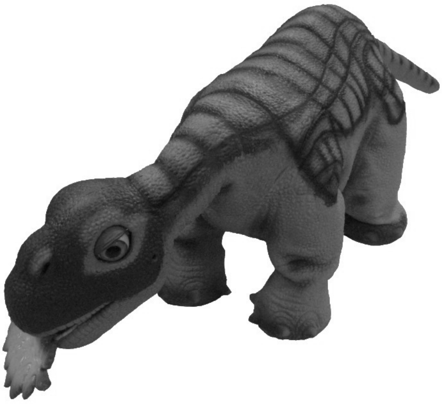


Figure 3. Pleo robot. (License: Creative Commons Attribution-Share Alike 3.0 Unported, author: Jiuguang Wang, source: https://commons.wikimedia.org/wiki/File:Pleo_robot.jpg)

Pleo (Figure 3) depicted a dinosaur and it was sold as a toy robot for children. They modeled it after a week-old baby Camarasaurus to embed all sensors and motors since this dinosaur has a big body and a relatively large head. Pleo could learn from its experiences and environment and the children need to feed him with toy food to be matured (Barnes et al., 2017). Despite Pleo's aesthetics, the owners expected locomotion skill, but it could not walk. Sony AIBOs (Figure 4) were released and manufactured by Sony and they were designed for entertainment without any utility, similar to Pleo. Most AIBO models resembled dogs, but they incorporated other inspirations into this product line (e.g. lion-cubs, space explorer). Sony robots could be raised from an initial pup stage to a fully grown adult and they went through their development stages by user interactions (Fujita, 2000). As we can see, the animalistic design of these commercial examples lowered the people's expectations for robot intelligence (Zaman et al., 2018). This strategy can help companies to ease artificial intelligence development for their robots. Although some robots were shown from the market, they are still initial developments and they usually meet the early adopters' expectations (Edwards et al., 2019).

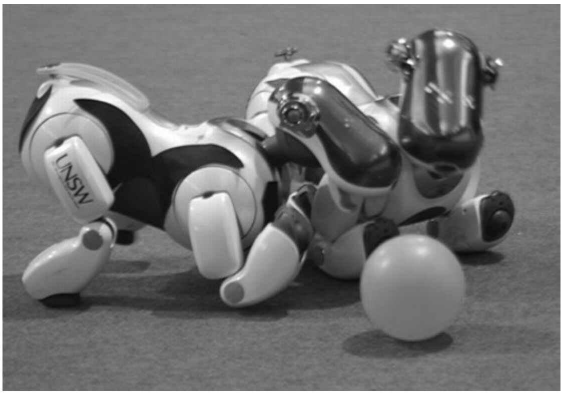


Figure 4. Sony ERS-7 robots in RoboCup competition. (License: Public Domain, author: Brad Hall, source: https://commons.wikimedia.org/wiki/File:RUNSWift_AIBOS.jpg)

Social robotics has a tight connection to human-robot interaction (HRI) research (Jung & Hinds, 2018) because the interaction is the primary medium to build social relationships between machines and humans (Miklósi et al., 2017). In this research field, it is a common intuition that participants accept robots better when they are capable of fluent interactions (Barnes et al., 2017). However, it was pointed out earlier, uncooperative behavior can exhibit social impressions (Horstmann & Krämer, 2020) thus the specific HRI situations will determine the required social skills (Lemaignan et al., 2017). The robots in hospitals and nursing homes (e.g. Paro) execute specific tasks to assist the elderly and improve the mental well-being of the patients (Nauta et al., 2019). On the other hand, the companion robots (e.g. Sony AIBO) entertains the owner as a primary function to create long-term emotional bonds (Björling et al., 2020). The target audience of the social robots is the ordinary people (D. Graaf, 2015), therefore, the built-in artificial intelligence in these robots is required to interpret human speech (Kennedy et al., 2017) and adapt to our behaviors (Rossi et al., 2017). Another challenge that the average users should not need to understand the technical details of a robot for everyday use to dissolve social robots in our society in the long run (Šabanović, 2010).

Social robots navigate in domestic environments designed for humans (Kostavelis et al., 2016) and their interaction style must follow the rules of our social behaviors (Pinter et al., 2015). People do not want to change their daily life because of the robots (Hiroi & Ito, 2013), they want to interact with them in the same way as they do with other humans. An ideal social robot communicates and interacts with human social terms (Talebpour & Martinoli, 2018) thus a human can empathize with it. This is the reason why HRI researchers take theories of human-human interactions to build social robots (P. Liu et al., 2016). Although this approach sounds easy, but even simple human social skills are hard to implement with a computer program limiting social robots' capabilities (M. M. de Graaf et al., 2019). Therefore, HRI research examines the social behaviors of robots in experiments to find optimal skill sets (Bajcsy et al., 2017) for interactions with humans including imitation (Doering et al., 2019), social learning (Zanatto et al., 2020) and maintaining relationships (Miklósi et al., 2017). The current prototypes in the research labs cannot present critical social skills to engage the users in natural interactions (Ferland et al., 2013). And as of today, the latest technologies are still constrained to develop robots with general artificial intelligence, so we cannot make such robots that are social in a real sense (Sünderhauf et al., 2018). A possible workaround is to program the robots to simulate social behaviors that people will perceive as social (Haring et al., 2013).

There are many aspects of social robots and human-robot interactions, but this dissertation focuses on the domestic robot companions at homes. Regarding this particular setting, the physical embodiment can facilitate social robots to become human companions (Deng et al., 2019). Appropriate aesthetics and robot design can maintain the right level of intelligence and social expectations (Ayesh et al., 2014). A robot with a pet appearance is treated differently than a humanoid robot since Mori showed (Mori, 1970) a relationship (uncanny valley) between the similarity of an object to an intelligent human or animal and the emotional response of a person to such object. His concept proposed that human-like objects can exhibit uncanny, eerie or unpleasant feelings in observers if their motions and appearances are not natural (von der Pütten & Krämer, 2012). While industrial and animal robots cause positive reactions, corpse-like robots (see Figure 5) invoke negative feelings. A well-done humanoid triggers more positive feelings than a toy robot (Fernaeus et al., 2010), but people are more tolerant of the mistakes of the latter (Mirnig et al., 2017) because of the lower initial expectations. However, negative emotions quickly emerge if a humanoid does not meet our requirements of human-level intelligence.

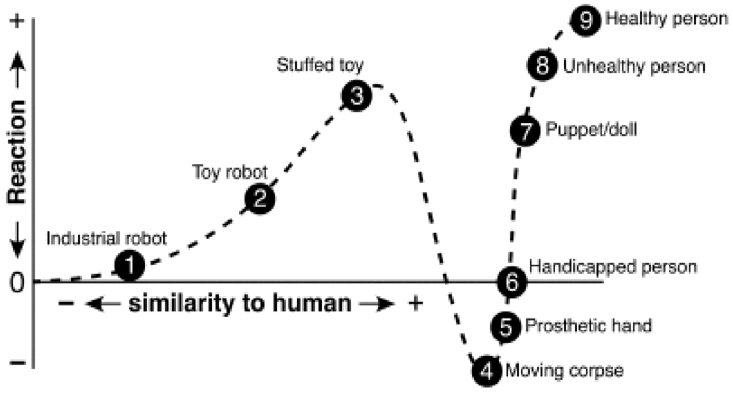


Figure 5. Uncanny valley illustration from (Mathur & Reichling, 2016). License: Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0).

If a robot wants to be a human companion, a suitable embodiment induces the right expectations from the owner (Deng et al., 2019), but the embedded social abilities are important as well for the ultimate robot acceptance (Heerink et al., 2006). Empathy is a key element to establish emotional bonds with humans (Leite, Pereira, et al., 2013) because the robot owners expect these social skills after the robot was advertised to be a companion. When a user finds out that his robot does not understand his expressed emotions and it acts without empathy, the user experience is ruined with a lowering acceptance level (Heerink et al., 2008). The user must feel that the robot cares about him and its actions can be trusted (Poulsen et al., 2018). An emphatic social robot recognizes emotions during communication (Zhang et al., 2013) and it can express its own feelings to the other party (Cameron et al., 2015). The humor (Mirnig et al., 2016) and politeness (Castro-González et al., 2016) are essential building blocks of the emphatic behaviors and the interactive conversations should be carried out smoothly to provide a consistent user experience (Heerink et al., 2010). Depending on the robot design (Ayesh et al., 2014), sound effects (Trovato et al., 2018), body gestures (Thimmesch-Gill et al., 2017), facial expressions (Churamani et al., 2017) and posture shifts (Obaid et al., 2016) are a rich set of modalities by which interactions can be augmented for better emphatic impressions.

As we could see in this section, social robots are a special kind of robot. They operate based on the same robot principles, but their purpose is bent towards the interactions with humans. Sony ERS-7 in this dissertation is a companion robot that is a subgenre of social robots. Their primary objective has no direct utility in the physical world, they are meant to be a companion for humans and improve their mental well-being. Despite the narrow focus of the companion robots, the technological realization of these robots is challenging because of the complexity and diverse requirements in the human interactions.

2.1.1 Recent Challenges in The Social Robotics Industry

Sci-fi movies and novels influence the public perception of robots thus there are incorrect preconceptions about the deliverable functionality with the latest technologies (Saffari et al., 2021) (Mubin et al., 2019). Thus social robotics companies must execute a market analysis before the product development in order to identify the beliefs of the target audience about social robots. When the marketing strategy is aligned with discovered customer expectations, the development roadmap will be realistic (Trivedi et al., 2018). The time to market period after the market research is crucial for the social robotics startups because it is the primary reason for innovation failures. When a robot is too long in development, the final product may not meet the target audience's original requirements anymore. Since the complexity of social robots is high (Seibt et al., 2021), unforeseeable obstacles can come up during the development. Therefore, the companies must maintain an active conversation with the community of potential early adopters or the backers on Kickstarter-Indiegogo webpages (Kindler et al., 2019). If it is not feasible to deliver the product on time, the company can attempt to realign the unrealistic expectations. Honest and transparent communication establishes the trust in the company (Lasschuijt, 2019) and its commitment to release the best possible product. However, the company should not attempt to fulfill a too broad target audience because the first customers or backers can be disappointed by the deliverables. These people will spread their initial negative experiences regardless of whether the actual product is niche.

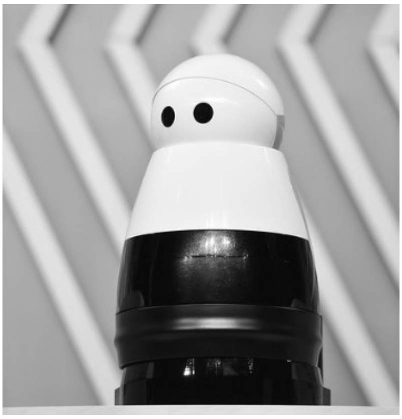


Figure 6. Kuri robot. (License: Creative Commons Attribution 2.0 Generic, author: Collision 2018 Conference, source: https://www.flickr.com/photos/collisionconf/27986269808)

Many social robotics companies fail caused by the inherent complexity of robots and the difficulties of entering the market. Kuri (Figure 6), an adorable home robot was developed by Mayfield Robotics, a subsidiary of Bosch (Ackerman, 2017). This robot was designed to play with children, respond to voice commands and patrol the home while recording videos and taking photos. The company did not plan a close interaction with the robot owners and 3rd party developers via an online forum to enable this community's potential. This robot was canceled right before its mobile application was launched and the first units were shipped for customers.

Cynthia Breazeal, a professor at MIT, launched an Indiegogo campaign in 2014 to introduce a new social robot. Jibo (Figure 7), a countertop robot, weighed nine pounds and it was 12 inches tall with a 6-inch base (Émond et al., 2020). The robot was stationary without wheels or legs for locomotion, but two spherical halves rotated on a three-axis motor system to animate his personality. The robot had a rectangular color display to show emotions and additional information. An application for handheld devices managed its settings and it could recognize up to 16 people by face and voice. The company ceased operations in 2018 because it could not secure further funding. One of the reasons, why Jibo failed, was that it did not offer more functionality than other devices already on the market (Bussgang & Snively, 2015). Competing products were smart speakers like the Amazon Echo, Google Home, smart home security cameras and virtual assistants in the handsets. These latter devices get us some information (e.g. weather forecast), perform common tasks (e.g. streaming music), but with greater flexibility and they have a low price compared to the robots.

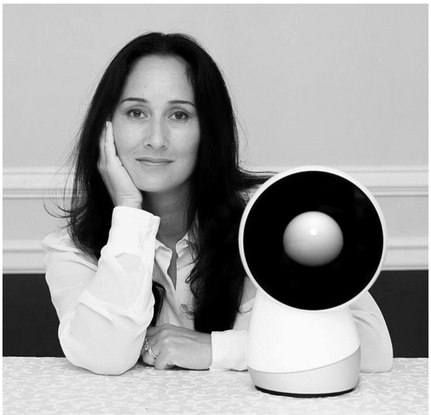


Figure 7. Jibo with its inventor Cynthia Breazeal. (License: Creative Commons Attribution-Share Alike 4.0 International, author: Cynthia Breazeal, source: https://commons.wikimedia.org/wiki/File:Cynthiabreazeal.jpg)

Jibo had an overpromising marketing video in the beginning and the product roadmap planning did not take into account the realities of the available technologies. For example, the conversational skills were not prototyped in the early phases which turned out to be a disaster. The available engines had too high network latency for a smooth conversation and they could not support other languages over English. The user experience was ruined. The company most likely did not execute initial market research about its target market segment's expectations. Without this survey, they did not define an achievable roadmap by their small team in a few years. The gap between the high initial customer expectations and the reality fired back later. Unfortunately, the company did not try to bend the expectations to a lower level by following the feedback of the early adopters of the Indiegogo community and their own development team. The startup promised to release a software development kit for 3rd party developers to write new skills for Jibo though it was never released. They set up a discussion forum (https://discuss.jibo.com) for the community where some Jibo team members made announcements about software updates and they addressed the questions of the Jibo owners. With this approach, they had the opportunity to incorporate direct feedback from the first customers.

A study from the Harvard Business School (HBS) (Bussgang & Snively, 2015) followed the company's internal roadmap and planning until 2016. In those years, the company seemed to be on the right track and had enough funding for operations. After reviewing the business goals and the internal planning in this study, several conclusions can be drawn. It was not publicly known, but the Jibo team was forced to launch the crowdfunding campaign because the investors of the upcoming Series A founding round wanted to have a proof of the targeted market size. The company considered the Indiegogo campaign as a direct measure to showcase the opportunities on the mass market for their social robot. They did not consider that the crowdfunding campaigns reach niche markets and early adopters thus the successes achieved on these platforms cannot guarantee a good launch on the mass market. They spared the efforts of proper market research to get a better understanding of their possibilities in the wild. This approach led to false believes about their opportunities inside the company.

The warnings of Series A investors did not ring cautious bells in the management when they asked about the differentiation between Jibo and smart assistants (e.g. Siri, Amazon Echo). What does justify the high price of Jibo over Siri or Echo? The HBS study clearly shows that the business management had no real idea about a social robot's additional values. They planned to release Jibo as a new app store for 3rd party developers to build new applications (Skills) for the platform. The company lacked a vision of how Jibo will build an emotional connection with some core skills to the customers. Without these capabilities, Jibo is a simple computer interface to the usual services accessible on mobile phones and smart speakers. Another overlooked problem was their imagination of what a 3rd party developer can build for Jibo with the current technologies. For example, they envisioned a remote pet watcher skill for Jibo which is still a challenging problem in 2020 after the vast advancement in deep learning since Jibo went defunct. And the AI skills for a robot require special knowledge from the software developers (deep learning, computer vision). It is not reasonable to expect production-ready AI features from 3rd party developers who used to program mobile applications for Android or Apple app stores. Furthermore, the business planning did not treat the promotional video for the Indiegogo campaign as a direct promise to release all presented features at the initial launch of the product. They regarded this video as a showcase demo to show the potential in the platform. They assumed that the 3rd party developers would build their promised Skills while the company will create only the essential core AI features. They did not understand that the crowdfunding sites' backers take the promotional videos as a promise for the shipped perks. The backers are very disappointed when they find out that a company never wanted to implement the presented features. This situation is dangerous for any young company with a niche product since these early adopters' opinions will affect the purchase decisions of future customers. The corporate strategy of Jibo can backfire for any social robotics startup because it is impossible to lower the backer expectations to a rational level until the product launch if the initial promises were already unrealistic.

Although Jibo and Kuri were different robot designs, these recent failures have a common issue. Namely, these robots could not be successful products because they did not have a compelling use case to attract customers. Their stories underline the importance of balancing between the promises and the reality of the latest technologies. The expectations of the messes are fueled by science fiction and people are unaware of the technology limitations. For example, a frequently cited robot, Roomba is in a different market (Jones, 2006) because it was built for a single task (cleaning). Therefore, people do not expect more than cleaning the floor efficiently. Roomba does not have inherent cuteness in the aesthetics which would imply additional feelings and expectations toward him as a companion. The robots captured our imaginations through science fiction for decades and it caused sky-high consumer expectations for autonomous robots. By these reasons, it is inevitable to identify a target niche community for early product planning and it must be included in the business plans of any future social robotics company. A misfortune was for Jibo that Amazon Alexa, Siri and other conversational agents entered the market after the company already had business plans, but it was not adapted to these new competitors. The competition was unequal. A startup attempted to develop a complex product without connected services while multinational companies developed the agents. These companies have overwhelming financial and human resources to integrate the agents into their existing ecosystems of digital services. An important lesson for the future social robotics companies is identifying the latest trends on the market and being ready to integrate digital services from other companies to satisfy the customers.

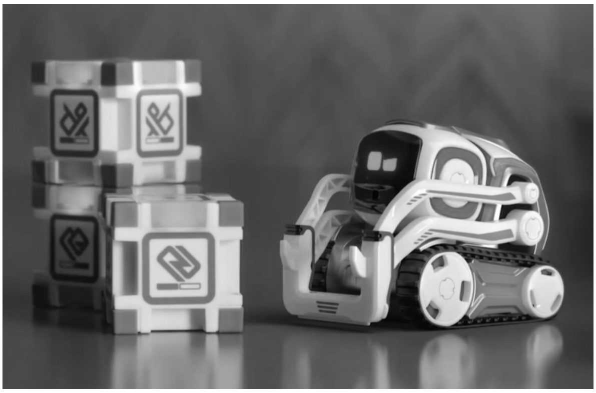


Figure 8. Cozmo robot. (License: Creative Commons Attribution, author: EP Daily and Dailymotion, source: https://www.dailymotion.com/video/x5y8hxt)

Cozmo (Figure 8) and Vector robots, originally made by Anki, have been successful social robots on the mass market. These robots do not appear to have human-level intelligence, rather than, they resemble a small, cute robot with a silly, childish personality (Chan et al., 2021). This approach implies that these robots can make mistakes and people forgive easily because these actions fit into their personalities. The clever design lowered the initial customer expectations and these robots were also priced to a fraction of Jibo. These appealing features brought success to Anki and it sold millions of units over the years. The community involvement was correctly implemented by Anki as well. They run an internet forum to connect with their customers (https://forums.anki.com), get feedback and encourage contributions. They released a software development kit to program Cozmo and the successful projects of 3rd party developers were featured inside the official application for smartphones.

Despite the fact that all strategies were executed properly to show off a great example in the robotics industry, Anki went bankrupt without precursors in 2019. Anki had nearly $100 million in revenue in 2017, they seemed to find a sweet spot with a sophisticated, affordable robotic toy. But the robotics development is not in line with the ordinary products, the hardware is expensive and the software must be developed continuously. Once a consumer buys a robot, he will have high expectations because of the costly purchase. However, long-term engagement requires upcoming software updates with new content. Since the cashflow stops from the customer after acquiring the robot, there is no sustainable source to finance software development. These special circumstances in this industry call for unique solutions like monthly subscriptions for software updates. After Anki went defunct, Digital Dream Labs bought all assets of that company to continue the robotics business. The new company learned from the mistake of Anki and they announced a subscription-based model for software updates of their Vector robot from October 2020 afterward. Sony went on a similar path after launching the newest Sony AIBO model (ERS-1000) in 2018 and the firmware updates come by a monthly subscription service.

These examples from the social robotics industry showed that it is challenging to launch a successful robot to the mass market. These products require careful planning in the prototyping phase and it is crucial to understand the target market segment. The communication to the potential customers is vital from the early stages to refine the roadmap iteratively and incorporate the unforeseeable, latest technologies until the market launch. The subscription models for software updates might be an inevitable building block in future strategies while involving the robot owner community is vital for product development. This latter is the focus of this doctoral thesis.

2.2 Communities in Open-Source Development

The open-source software (OSS) emerged into a base asset for multiple industries, including telecommunications, robotics and cloud services. Since the companies become dependent on larger OSS stacks, they hired full-time employees to work on these free projects.

Open-source projects are usually shown as a distributed team of volunteers that builds a community to execute their common goal. Although this definition is true in many cases, long-standing projects receive considerable corporate support. The companies have different reasons to allocate human resources to OSS and some expect external contributions into their software to gain economic benefits. They rarely realize that contributing back to the upstream projects teaches by experiences how to use the OSS more efficiently in their corporate environment (Nagle, 2017). Nagle found, the usage of OSS in contributing companies resulted in higher overall productivity and a competitive edge over the competitors. When companies dedicate employees to an open-source project for a longer period, it improves the stability of the OSS project and increases the company reputation. Nowadays, technology companies earn good credibility by contributing to OSS projects and this strategy shows the firm's social responsibility for outsiders.

Usually, the open-source projects depend on a few *core developers* who contribute the most significant parts of the code, maintain and administrate the web services. These people are less than 25 % of the contributors (Dinh-Trong & Bieman, 2005), but most researches analyzed these developers. The majority of the community contributes infrequently, bthese contributors are called *peripheral developers*. Although the one-time contributors are common, some peripheral developers participate for the longevity of the project. (Barcomb et al., 2018) divided these developers further by the frequency of their contributions. *Habitual volunteers* make either frequent contributions (10 or more in a year) or their participation lasts for a sustained duration (2 or more in each month for a half year). *Episodic volunteers* are peripheral who contribute less than the habitual, for example, a few times over the years. These latter contributions are seldom, but their committers return time-to-time and their retention is desired because they are over their initial learning curve. Apart from the volunteers, the contributing paid developers are remunerated by a company or a foundation and there is less need to motivate them. The literature reviewed mainly the code contributions in the past (Carillo et al., 2017), but translations, management of web services, documentation and artwork are all important for the long-term success and the peripheral developers can contribute to these tasks. This potential is significant for this dissertation because the AIBO community consists of mainly non-technical people.

The supporting technologies and services around the OSS projects enhanced in the past decade. Git became dominant for version control and the markdown language democratized the documentation writing for non-technical people. Several homepages provided hassle-free project management with git hosting, wiki pages, bug tracking and code reviews (e.g. github.com, gitlab.com). The simplified access to the contribution process attracted more developers than ever before.

Github is a web service for version control thus the typical hosted projects contain program code, but some are bookmark collections or textbooks. This service is used around the globe, 37 million users and 57 million repositories are registered. The peripheral developers made 7% of Github projects' contributions in 2012 (Gousios et al., 2014), but their number increased to almost 50% after a few years (Pinto et al., 2016). The peripheral contributions were believed to be grammatical corrections in the documentation or new translations for applications. Nevertheless, the more in-depth analysis showed that the majority was code contribution. One-third of them fixed bugs, one fifth submitted new features and 9 % was related to code refactoring. These contributions were driven by personal demand and they were found beneficial by the core developers although it involved more administration and code review. Most peripheral developers on Github were habitual volunteers since most of them (63 %) made at least one contribution every month. However, they were episodic volunteers from the OSS projects' point of view since half of their contributions were one-off. These observations portrayed a persona for a typical peripheral developer on Github. He uses the products of multiple open-source projects on his computer at work or home and he submits smaller fixes to the original project when small problems are encountered. These quick contributions were supported by the low barrier to submit changes on Github which is a reason why it became the Facebook of open-source projects. However, there are differences between the projects how many contributions they receive. The OSS projects written in static typed high-level languages received 2-3 times more peripheral contributions than languages with dynamic type checking. This phenomenon is important for the author's community efforts because the target Sony AIBO robots can be programmed in a statically typed language (C++).

Pinto et al. (Pinto et al., 2016) executed two surveys on Github to understand how the peripheral developers think about their contributions and why they did not contribute more frequently. The personal needs were cited to be the greatest motivation for submitting a change. When a typical peripheral developer is blocked by an issue, it will be solved and submitted on Github because the easy contribution process facilitates these small fixes. Other motivations were to contribute back to the community, build a reputation and improve certain projects. Although the episodic developers have good motives to contribute to open-source software, their relationship to the projects do not evolve to habitual contributions. Half of the episodic developers mentioned the lack of time as a primary reason to hold them back from the next step. Much closed-source software is based on OSS and their paid developers contribute the necessary changes back to the community, but they do not want to sacrifice free time to continue the contribution. The unpaid episodic volunteers are not remunerated for their OSS contributions and they do not invest more time because their personal motivation is not high enough without additional incentives. Another explanation is the lack of deeper motivation since the original reason for the contribution was to solve a simple problem and move on. Some episodic volunteers had inadequate skills or knowledge to prepare bigger contributions thus they preferred low efforts to solve small problems. The overall impression of these surveys was that the peripheral contributions are positive for both the contributors and the core developers. The new eyes help to discover unnoticed bugs and establish continuous code improvements. The minor drawbacks were the spent time by core developers on review and the risk of unmaintained code from contributions in the future.

Since the AIBO community consists of mainly non-technical people, it is not reasonable to expect they become core developers, but a habitual or episodic role is more likely. The peripheral developers are quite common in the OSS world and a large portion of them contributes small changes. This aspect of the collaborative OSS development can be utilized for social robots and the small improvements can be integrated in the robotics software with careful planning and testing. The positive feedback of the project maintainers about the OSS peripheral developers depicts fruitful outcomes for this dissertation. However, the author does not have a project budget to pay remuneration for the contributions. When the AIBO community members contributed to this thesis project, they were driven by their personal motives and engagement which would be a real success compared to paid contributions.

Barcomb et al. (Barcomb et al., 2018) interviewed open-source developers to suggest strategies for the core developers on managing the peripheral contributors. Such tips can enhance developer retention and result in more habitual contributions. The motivations of the contributors are a mixture of altruism and self-centered motives. Hyde et al. (Hyde et al., 2016) shown that both altruistic and self-centered motives were equally present among the newcomers and habitual contributors. Extrinsic motives like remuneration were not crucial to retain the contributors in the long run (Krishnamurthy et al., 2016). Those interviewed newcomers remained with a project who had intrinsic or altruistic motivations. The peripheral contributions were driven by a momentary relationship between the developer and the project, the general feeling of the developer towards the OSS projects was not important in these decisions. The volunteers contributed to the projects without remuneration because they had an interest and they enjoyed. However, their commitments were restricted by their daily job, family affairs and the available leisure time. Some projects employed general notifications for participation when there was a need for certain contributions. Though this method was not created for the peripheral developers, but it was effective to reach these developers who were interested in only specific tasks.

The analyzed OSS projects lacked any practices to retain peripheral developers though predefined small tasks (e.g. bugs, translations) were suggested for newcomers to encourage the initial peripheral contributions. Although the source code contributions are thanked in some automated way, the non-code submissions and bug reports are not honored anywhere. The predefined tasks facilitate a quick and useful contribution to prevent discouragement by missing technical skills. The easy, standardized contribution process on Github simplified the submission and review processes while it eliminates the learning curve of a project-specific web interface. The socially motivated contributors can be guided on thematic events of the project to build personal connections and overcome technical difficulties (e.g. Akademy event for KDE project). If such an event is not reachable for a new contributor, active communication channels (chat, mailing list) can ease the initial steps. However, good language skills are vital because the natural communication language is English inside the international communities and the available translations of the documentation are fairly limited in many OSS projects.

The community feeling is important because it incorporates efficacy, support and responsibility. When a developer receives pressure or support from other project members, these social norms can positively affect finishing a contribution. These encouragements are successful for the novice contributors to become habitual (Hyde et al., 2016). The open-source developers experience affinity for the project they work for (Bagozzi & Dholakia, 2006), even the episodic contributions produce positive thinking in the developers. However, the developers, who experience this pleasure, are more likely to continue their contributions. The episodic developers tend to feel less attachment to an OSS project, but the habitual developers cited the community feeling why they contributed later again. The satisfaction is triggered when the initial expectations are in line with the returned community feedback. And similar to the traditional volunteering, satisfaction is the best indicator to identify the future habitual contributors (Wu et al., 2007). The peripheral developers are satisfied with their OSS contributions when they feel appreciated and help others in the community. However, in bigger projects, the open-source software is released in regular time periods thus it is difficult to count on peripheral developers with unpredictable schedules.

2.2.1 Strategies for Peripheral developers

The open-source projects do not employ strategies to identify and manage the possible peripheral developers, all novice contributors experience the same treatment. A sophisticated action plan can identify a newbie's goals, the fitting tasks, supporting the initial efforts and measuring the progress. A cost-benefit analysis can justify if the peripheral contributors bring value to the project in case of a corporate collaborative project.

First, the community must be analyzed to understand the most effective ways to motivate members and get know the different skill sets of the peripheral contributors. *Gurus* do not need to deal with any learning curve because they are familiar with the project environment. Novices require guidance to the contribution process and a mentor to solve the obstacles. Gurus and novices are capable of developing different tasks with their skills. As peripheral contributors can work on a wide range of problems over the traditional code contributions, it is a matter of good scheduling to involve these people efficiently. A project hosting service must be chosen with quick registration and a straightforward web interface to ensure the easy contribution process. Good examples are Github, Gitlab, Bitbucket or Launchpad. When the newcomers have strong social motives, they can get good impulses after inviting them to a project gathering (e.g. Akademy by KDE project). The recognition of non-code contributions must be included in the project processes to increase the peripheral contributors' satisfaction. Later, follow up announcements for targeted tasks effectively encourage the peripheral developers to return for a new contribution.

2.3 Open-Source and Community Involvement in a Social Robotics Company

The robots interact with people and they must interpret the human actions properly to execute a correct response. Solving this situation is rather challenging with current technologies. The majority of the social robots have been developed by small companies (Cozmo by Anki, Jibo, Sophia by Hanson Robotics etc.) which have limited human resources for software development compared to the complexities in the human-robot interactions. None of these companies grew out of the early market to the mass market although some products (Cozmo, Sony AIBO) had notable popularity.

The main difficulty for the social robotics companies is the expensive joint hardware and software development for a new robot. If the production costs cannot be lowered to an acceptable level for the mass market, the robots will always stay in their niche market. Next to the rising research and development costs, the startups have constant pressure from the investors to fulfill the upcoming seed rounds and the short time-to-market cycles.

These risks are originated in the immature founding technologies and the gap between the successful research experiments in a lab and the real-world expectations for the production-level quality. The startups usually introduce their robots to the early-adopters via Indiegogo or Kickstarter campaigns these years which are considered the natural online platforms to reach niche markets.

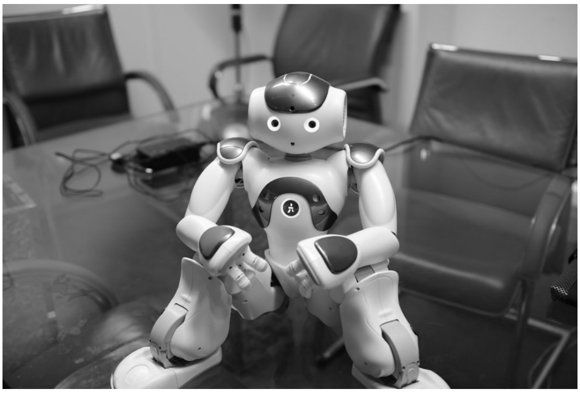


Figure 9. Nao robot. (License: Creative Commons Attribution 2.0 Generic, author: Stephen Chin, source: https://www.flickr.com/photos/steveonjava/8170243076/)

The author interviewed four people (a software developer, a researcher and two managers) from Aldebaran Robotics in 2016 (before the acquisition by SoftBank), which developed the Nao (Figure 9) and Pepper robots. Their answers were collected by email exchanges and they gave insight into how the community and open-source tools are utilized in one of the world's biggest robotics companies. At the time of the interviews, around 400 people were employed inside the company and 60 % of the workforce was either a software developer, tester or hardware engineer. The company followed modern, incremental and agile software development processes to build artificial intelligence for their robots. New source codes by the employees were reviewed and validated by unit, modular, functional and manual tests before they were merged into the codebase. The testing is an important step before the software changes become part of the latest software release thus the company taken the quality control seriously with separated testing-integration teams inside the organization. Once any stakeholder in the process discovered a serious bug in the release candidate, the new software update was blocked. A common problem in software testing for robots that the expensive hardware can be damaged hence it is valuable to move some verifications in simulated environments. Aldebaran Robotics used both strategies, they had simulated and manual tests on the robot.

The open-source involvement can help the internal software development in companies to reach their goals quicker. Aldebaran Robotics used open-source tools and software with success. They found the OSS beneficial and effective as the bleeding edge technologies require the usage of open-source software to provide the Aldebaran's plus value in the competition. Some of their own software components were opened and they were publicly available on Github (https://github.com/aldebaran). They defined processes to get user contributions into their software to involve 3rd party developers and foster further partnerships with other technology companies.

The author evaluated the Github contributions to the software components of Aldebaran Robotics. 18 low-level software components and coding tutorials received 227 contributions over 9 years. The issue reports (67.84 %) and the actual submissions (32.16 %) were the two types of contributions. 3.8 % of the issues and 16.43 % of the submissions were documentation related while the rest was source code changes. This result is similar to a meta-analysis on Github (Pinto et al., 2016) in the literature where the fraction of the contributions was documentation related and the majority was source code changes. The author had no information about the number of paid core developers by Aldebaran Robotics. However, 93 3rd party programmers contributed to the software although none of them were core developers, all were peripheral. The majority of this developer community on Github was episodic with a few contributions, but 6 habitual developers (6.4 %) made more frequent submissions over the years. This ratio between the episodic and habitual developers chimes in with the literature regarding Github where most peripheral developers were episodic. The company received many contributions through this channel, they found this practice good and planned to open up more software components.

Aldebaran Robotics involved a community on Github to get bug reports and code contributions, but this web service reaches the highly technical community members. The company opened an internet forum to widen the community and discuss all matters about their robots. There were 2 full-time employees to deal with the community relations daily and some engineers joined the discussions on an irregular basis. The forum was part of the company website and none of the community members were involved in the forum administration to offload the community team. However, depending on the community members' involvement, they could communicate with managers and even directors in some specific topics. Their public Github repositories did not contain any high-level software and they intended the forum as an interface to engage the creation of new body animations and dialogs for their robots. Though the development ideas and feature requests were discussed on the forums, the community team focused more on the direct feedback of actual 3rd party developers who created new content for the robot. This feedback was given to the internal engineering teams which were in charge of the core development for the robots. The company never expected direct software contributions from the community members on the forum to the official high-level robot software unlike to the core software on Github. The forum members were mainly professional developers, university students and researchers who joined to get answers for their technical questions about the Nao robot. Most active members on the forum owned a robot and they were professional developers, however, the forum attracted non-technical people over the years and these members were interested in discussing robotics-related issues in general.

As we can see, the Aldebaran Robotics had a clear vision to open up software components for 3rd party developers to engage external contributions and collaboration with other companies. They uploaded software packages to Github, defined processes to get contributions through this channel and dedicated paid employees to deal with this community. They targeted technical contributions to their lower-level components on Github and their forum had the vision to encourage the creation of high-level behaviors and motions for their Nao robot. These activities attracted people and an online community was formed around their products, the overall external contributions were valuable for the company and planned to extend these online activities. It is worth noting that further engagement methods were mentioned in Section 2.2.1 which were lacked in the Aldebaran's strategy. They did not execute cost-benefit analysis to get an objective metric about the return-of-investment benefits for the company while the dedicated community team and the irregular commitments of internal engineers are recurring costs for the company.

Although the online activities were well-defined, the occasional incentives were missing for 3rd party developers and they did not organize community events to meet with the community members in person. The external developers could get a Nao robot for a discounted price in the frames of the Nao Developer Program, but there were no more incentives after joining the program. This developer program targeted gurus while the students (novice developers) could access Nao robot in their university labs. Though online communication was established with the community, the company was not proactive in encouraging certain development directions with the community members. To summarize this analysis, Aldebaran Robotics showed a good example of how to plan and execute a community involvement for artificial intelligence development, but there are still areas to improve for social robotics companies in the future.

2.4 SUMMARY

The social robots and their industry were reviewed in the first part of this chapter. Some hidden obstacles in AI development were revealed despite our advanced technologies and how they are often realized by companies too late. Afterwards, the essential basics of collaborative open-source development, the roles of the core and peripheral developers were shown.

The possible solutions for AI development were discussed and the collaborative development was proposed as a viable alternative.

The next chapter describes a case study application where the author executed an open-source software project. He was part of the online discussions and several robot owners from this community were involved in the artificial intelligence development for their social robot. This example project provided additional experiences on how social robotics companies can release more sustainable robot products to the retail market.

3 Collaborative Development for a Social Robot

The collaborative AI development for social robotics is an unexplored research area in robotics without any previous publications in this topic to the author's best knowledge. There is no established research methodology, and the industry's existing practices are limited (see Chapter 2.3). Therefore, the author executed an open-source project and involved long-term robot owners in active development. This chapter describes how AiBO+ project was started, carried out for years and what kind of experiences were learned to enhance future robotics projects.

3.1 THE PROJECT START

The author intended to develop an AI engine for a mobile robot which runs on a battery, it can navigate in a room autonomously and it is equipped with sensors to perceive the environment. The commercial market of mobile robots was scarce in the early 2000s because the underlying technologies were not good enough. The onboard battery capacity limited the wheeled and legged robots to 30-120 minutes-long operation. The embedded processors were far below the computational capacity of the modern ARM chips. The lack of a standardized robotics platform and the missing intelligent cloud services (e.g. facial recognition) made the artificial intelligence development hard. The author surveyed the available robots on the market and chosen the AIBO robots despite Sony already stopped the commercial sale of these robots caused by drastic financial cuts in 2006. The AIBO robots are quadruped hence their locomotion is more complicated than wheeled, but easier than bipedal robots. All AIBO models had various sensors to sense the environment and a low-resolution color camera to capture the scene before the robot. The battery runs the robot with one charge for 1.5-3 hours depending on the activities and low-level sensor access was granted by Sony to develop custom software for these robots.

Other robots were not close to the capabilities of the Sony AIBOs in the 2000s. Genibo was a Korean pet robot with a similar concept to AIBO, but its artificial intelligence was far behind and Dasatech never released a software development environment (SDE for Genibo. i-Cybie was a low-cost robot dog from a Hong Kong company that had basic hardware and no programming interface. Pleo was an animatronic pet dinosaur toy which had limited walking capabilities, low-quality build, short battery life and no SDE. Lego Mindstorms is an interesting modular concept to build robots, but a color camera is not available in the kit and the onboard processing power is constrained.

3.2 Sony AIBO

The Sony Computer Science Laboratories (CSL) was founded in 1988 to focus on fundamental research in computer systems. The organization shifted the research efforts over time to explore applications in systems biology, econophysics and artificial intelligence. Dr. Toshitada Doi and Masahiro Fujita worked together in Sony CSL to build the early prototypes of Sony AIBO as early as in 1993. The technology was limited at that time, therefore, they designed a robot for entertainment purposes. The prototyping was so successful that the initial model was introduced to the public in 1999.

The abbreviation AIBO comes from **A**rtificial **I**ntelligence ro**BO**t and the word means *companion* or *friend* in Japanese. The major concept of these robots was the continuous interaction with the owner and creating an emotional bond. The robot was shipped with a newborn personality and the robotic pet had to be raised to an adult personality by interactions. The first generation (ERS-11x) was very successful in 1999, they were sold out in record amount. The next generations (ERS-2x0, ERS-33x) were released in 2000-2001, the aesthetics and the hardware were improved in these models. The fourth-generation was introduced with the Sony ERS-7. This model featured the best hardware and built-in wireless connectivity. This dissertation was based on this model. Sony started to sell the ERS-1000 model in 2018, but this new model does not offer an open SDE with low-level sensor access unlike the earlier models.

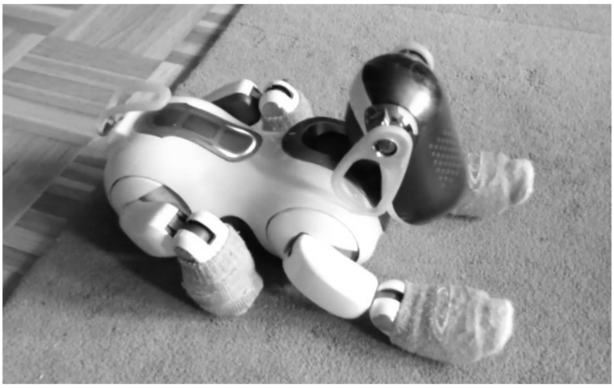


Figure 10. Sony ERS-7 robot. (Own work)

3.2.1 Hardware Details

The robots are basically mini computers at their core, but they have cameras and sensors to sense the environment and they can utilize motors to move around. This short section details the hardware capabilities of Sony ERS-7, the robot used in this dissertation.

The Sony ERS-7 (Figure 10) had an RM-7000 processor which was based on MIPS architecture and clocked at 576 MHz. The applications were stored in a removable Sony Memorystick with up to 128 MB storage space and 64 MB RAM was available to execute artificial intelligence programs. The robot walked by four legs with 3 degrees of freedom and paw sensors measured feedback of the ground contact forces. The head could turn around to watch the surroundings with a CMOS camera and certain objects could be grabbed with a mouth. Two infrared sensors in the front of the robot looked forward for near obstacles and an accelerometer with a 120 Hz sampling rate in the torso estimated the body movements in the 3D space. Stereo microphones in the ears listened to the audio cues during human-robot interactions and a miniature speaker played back sound effects. LEDs on the face and the body gave visual feedback to the human observer and petting can be sensed by touch sensors on the robot body.

3.2.2 Software Development Environment

The AIBO robots were bundled with artificial intelligence software from Sony to entertain the owner with a certain personality, walk around and interact with toys. The ERS-7 software was called Mind and it included all AI features shipped for previous models. However, the official software was not accessible for the 3rd party software developers to manipulate or write extensions. All AIBO products run under the Aperios proprietary operating system to implement real-time capabilities on resource-scarce embedded systems. A programming interface (Open-R) to the operating system was opened for robot specialists in 2004 to facilitate the social robot industry's growth. With this option, new software could be developed for the Sony ERS-7 model and distributed for commercial and non-commercial purposes without any license fees. This Open-R SDE included a C++ programming interface to the low-level hardware (sensors, camera, motors). The main disadvantage of Open-R was the lack of AI skills. Sony did not include any existing Mind functionality into the Open-R SDE, the artificial intelligence development must be started from scratch. Hence the author of this thesis is a professional C++ coder and passionate about robots, he devoted some years to write new software for this robot and engage other community members to contribute to AiBO+.

The software development for robots are usually based on Android or Linux, but the Aperios operating system lacked a similar environment and some standard C library functions. The developed program must be run on the robot perfectly otherwise the robot suddenly stopped and shut down in case of any crash or large memory leak. The official Sony SDE supported C++ language that needs some standard software to turn the program codes into binary files to be run on the robot. This software is called a toolchain and it contains a compiler program and related tools. The Sony toolchain relied on three key components besides the Open-R system libraries: a compiler (gcc 3.3), linker tools (binutils 2.15) and a minimal C system library (newlib 1.15). These components were already outdated in 2006 when the sales of AIBO robots were stopped. The author updated these C++ compiler tools to gcc 5.4 and binutils 2.24. This upgrade brought the latest C++ standard features (C++11, C++14) and open-source software for Sony ERS-7 (Kertész, 2013). The enhanced software environment could use Boost[1] for data serialization, tiny-dnn[2] for deep learning and OpenCV[3] for image processing. This robot has a little RAM (64 MB) for executing the AI algorithms and storing all data, therefore, it was essential to minimize the memory consumption. The ported C++ compilers implemented new code optimizations to lower the program size and 25 % of final reduction was achieved by optimization flags and tunings for the MIPS processor on Sony ERS-7.

3.3 Understanding the Target Audience

After the author acquired basic technical skills to develop programs for the Sony robot, it had to be decided which aspects of the AI should be implemented first due to the time constraints. The primary driver of this software development was to engage community members in contributing to the AiBO+ project. As it was noted in Section 2.2.1, a community must be

[1] https://www.boost.org

[2] https://github.com/tiny-dnn/tiny-dnn

[3] https://opencv.org

surveyed before a collaborative project is started since the community size, the member expectations and their skills must be known for planning purposes. Because of these reasons, the author joined an internet forum where the AIBO owners met online to discuss all matters about their robots. A questionnaire was run among these forum members and the answers were evaluated in Publication I and II. The evaluation details can be read in the following short summaries.

A Short Introduction to The Questionnaire

The data collection from participants is usually done by filling out a questionnaire in social sciences. The questionnaire in this dissertation used typical question types and asked for answers in free text and Likert-type items. A Likert-type item can receive a numerical response between e.g. 1-7 where the numbers represent a scale, 1 for strongly disagree and 7 for strongly agree. The free text answers are useful because a participant can express his opinion without constraints while the Likert scale makes the answers available for quantitative analysis. Three primary methods were used to analyze the quantitative answers. Null hypotheses were defined according to the research questions and statistical tests verified if these null hypotheses were accepted or rejected. If the quantitative variable of the hypothesis had two categories, the Mann-Withney test was used, when it had more categories, the Kruskal-Wallis test was employed.

Reference

Kertész, C., & Turunen, M. (2018). Exploratory Analysis of Sony AIBO Users. *AI & Society*, 1–14. doi:10.1007/s00146-018-0818-8

Objective and Method

This exploratory study examined the long-term owners (heavy users) of the Sony AIBO robots to discover their robot acceptance phase and expectations for a hypothetical software upgrade. They filled a questionnaire whose introduction part asked basic questions (gender, age, home location, profession) and about the robot ownership (length, usage frequency, model preference). The main questions of the survey were composed of 9-point Likert-type items to get answers for quantitative analysis. The questions touched the expectations from a new software update, wishes for new skills, connectivity options, autonomous behaviors and future user contributions. 78 participants answered, 57 males and 19 females, mainly from developed countries. They were separated for Westerners and Japanese to examine our stereotype that Japanese people have a special, accepting relationship with robots compared to other countries. This stereotype was studied in the literature to find culture-based explanations (Kaplan, 2004) (Šabanović, 2014). The survey responses were analyzed to answer the following research questions:

- How does the length of ownership affect the perception of the robot?

- Is there any significant difference between Westerners and Japanese people?
- Does the age change the users' opinion?
- Does gender make any difference?
- How much do the heavy users feel inclined to contribute to the Sony AIBO software?
- Which Sony AIBO models are preferred by the heavy users?

Results

The Likert-type items were grouped into four subscales. There are statistical methods to verify the consistency of these groups after the answers were collected. Their Cronbach's α coefficients were above 0.8 which refers to good trust in the overall reliability of the grouping. The exploratory factory analysis is another method to uncover the coherency inside the subscales. All Likert-type answers were considered as independent variables, they were formed into new groups (factors) with this unsupervised algorithm and the detected factors were almost completely identical to the subscales. The Kaiser-Meyer-Olkin Measure of Sampling Adequacy was 1, Bartlett's Test of Sphericity was significant under $p < 0.001$ for approximate of Chi-Square 3649.37 thus the measured variables were not normally distributed, but skewed. 12 Likert-type items had eigenvalue over 1.00 and they expressed 78.17% of the total variance. After the overall reliability of the results was confirmed, an analysis was carried out to answer the research questions.

The subscale answers were evaluated to see general tendencies. It was surprising that people attributed life-like properties to the robot after years of ownership. The participants preferred interaction with the robot and more autonomous skills instead of the repetitive entertainment behaviors of the factory software. They wished their latest gadgets could connect to Sony robots in order to check the robot state or emotions, but they were not interested in controlling them remotely.

Four null hypotheses were defined to explore the first four research questions regarding the gender, age, culture and length of ownership variables. If a variable had 2 categories, the Likert-type items were evaluated with the Mann-Withney test and the Kruskal-Wallis test was performed for more categories. After these non-parametric tests, the null hypotheses were either accepted or rejected. Significant differences in the medians of the Likert-type items were discussed when their p-value was low.

Gender null hypothesis (H1): *The male heavy users see a social robot as a machine and the female as a companion*.

The survey had 19 female participants and 57 males. These two categories were tested with the Mann-Whitney test and the results confirmed the common stereotypes, the women tended to be more emotional in their ratings while the men were technology-minded. This hypothesis was accepted.

Age null hypothesis (H2): *The younger heavy users are more technology-minded while the elder look the social robots as a companion*.

Three categories were defined to analyze this hypothesis with the Kruskal-Wallis test. 11 participants were under 25 years, 24 between 25-40 years and 43 were over 40 years. The results were against the null hypothesis hence it was rejected. The older people did not perceive Sony AIBO as a companion to a greater extent and the younger generations were not more eager about the technology side of these robots.

Culture null hypothesis (H3): *The Japanese heavy users do not rate their robots more positively than Westerners.*

The cultural background was examined with this hypothesis on two categories. The Westerners filled an English version of the questionnaire while the Japanese completed in their native language. Japanese found these robots more boring and they underrated their technical and emotional skills, therefore, the null hypothesis was accepted.

Length of ownership null hypothesis (H4): *The more years a heavy user owns a social robot without software updates the more robot acceptance decreases and he/she loses interest over time*.

Four categories were defined for the length of ownership: below 2 years, 2-5 years, 5-10 years and over 10 years. It was revealed that the consumer interest of the heavy users did not decline after years of usage. The users had constant anthropomorphic characterization regardless of the passing time, but the need for autonomous and social features increased after 5 years. This null hypothesis was rejected.

The questionnaire asked about the user contribution regarding the fifth research question. Almost two-thirds of the owners (59%) expressed willingness to make new content for the robot, about one-third (31%) refused and the rest was unsure. The Japanese owners wanted to contribute with 30% more chance than Westerners.

The Sony AIBO product line had several robot dogs with different capabilities. This article analyzed the model preference of heavy users. The favorite models were the most advanced ERS-7 and the ERS-2xx models with some autonomous and interaction skills. The participants did not appreciate the cheapest ERS-3xx models because they had the least software options and skills. The young people did not have any clear model preference and the Japanese always chosen only one model.

Discussion

According to common knowledge, women are more emotional than men and some earlier scientific evidences supported this stereotype (Brody, 1997) (Bradley et al., 2001) (Seidlitz & Diener, 1998). This expectation was reflected in H1 what strengthens the findings in (Scopelliti et al., 2005) with Westerners, but the literature had mixed results in Japanese society. Two surveys did not show a difference in the genders (Nomura et al., 2005) (Nomura et al., 2012), but the female Japanese students were more positive towards robots in (Nomura et al., 2006).

The H2 hypothesis reviewed the results from the age point of view. The younger generations among the participants were more positive towards the robots, but they were not more interested in their technology side than the elder. Previous researches found the same results with Australians (Zhan et al., 2016), Italian (Scopelliti et al., 2005) and American (Ezer, 2008) participants. However, two studies found the opposite in Japanese society (Nomura et al., 2005) (Nomura et al., 2012). Since the majority of the participants were from Western countries in this article, the sampling can explain why the current results are similar to the Western societies in the literature.

The cultural hypothesis (H3) explored the common belief that Japanese people love robots more than Westerners. Some studies (Bartneck et al., 2007) (Haring et al., 2014) indicated that this stereotype is not real and the Japanese are not more enthusiastic with robots than Westerners. On the contrary, they were more negative than Westerners in this research, similar to (Haring et al., 2015).

The H4 hypothesis found that the heavy users did not abandon their Sony AIBO robots after years of ownership although common sense suggests the opposite because of the decreasing utility value over time. A proposed reason for this outcome was the developed emotional attachment of the owners towards their robots.

The older generations were more likely to contribute to the robot software although the younger generations should be more familiar with the latest technologies. The Japanese participants were interested in collaborative development despite being more negative about robots in the H3 hypothesis.

The analysis of model preference revealed that the heavy users preferred the most intelligent robots with rich skills despite their higher prices.

Based on the results of the article, design recommendations were given for social robots to complement the past works (Leite, Martinho, et al., 2013) (M. M. de Graaf et al., 2016). The long-term ownership did not bias robot acceptance and all age groups were continuously present over the years. The heavy users appreciated their robot, but they desired the integration of the newest technologies into the ecosystem. The robots must show real intelligence to differentiate from normal machines. The robots must consider gender, age and culture in their communication.

The analysis was done with 78 participants which is a limitation since the Westerners were overrepresented. However, this study examined heavy users with years of experience in robot ownership which makes the results essential.

Reference

Kertész, C., & Turunen, M. (2017). What Can We Learn from the Long-Term Users of a Social Robot? In Proceedings of the 9th International Conference on Social Robotics, 657– 665. doi:10.1007/978-3-319-70022-9_65

Objective and Method

This study evaluated the second part of the answers to the questionnaire analyzed in Publication I. The past articles in the literature executed experiments with participants in weekly or monthly sessions where they interacted with a robot. However, the analyzed users in this article were lived with Sony robots for years together and they used them from time to time instead of laid in the storage room. The article explores the technical expectations of these heavy users and what kind of improvements do they expect to remain in the acceptance phase?

The questionnaire asked the gender, age, home location and profession in the first part to get basic information about the participants and feedback was gathered about their robot ownership including the duration, usage frequency and AIBO model preference. The following questions with 9-point Likert-type items were related to the perception of their robots and additional information could be written to optional text fields. Two questions were about desired skills and connectivity options as well as the free-form answers were analyzed to characterize the long-term user expectations.

Results

61 persons were recruited to fill the questionnaire from an English-speaking online AIBO forum and 17 Japanese from Facebook with a targeted ad

campaign. The gender distribution was 57 males and 19 females. The age distribution was healthy across all ranges. The participants kept their robots after the technology acceptance phase with a high retention rate. 20% of the participants had their robots for more than 10 years, 51% had between 2-10 years and 28% had for less than 2 years which is a high proportion of newcomers. Most owners had technology-related jobs e.g. engineers, software developers or technicians (Figure 11).

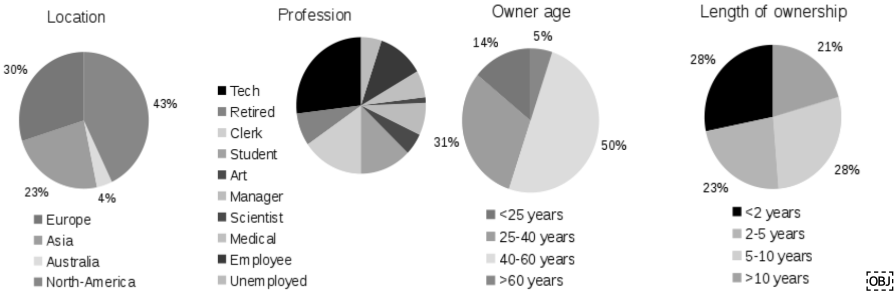


Figure 11. Location, profession, age distribution and length of ownership of Sony AIBO users among the participants.

A quantitative analysis examined the Likert-type questions about the technical expectations and the consistency of the answers was verified by Cronbach's α coefficients (0.91, 0.81). The participants had a low interest in enhancing the non-interactive entertainment features like dancing or tricks with toys. However, the human-robot interaction skills received a high interest like speech recognition or distinguishing humans. These results confirmed that the owners valued the emotional connection with their robots above the pure entertainment behaviors. Autonomous features got high ratings and the participants wanted to connect their robots with their mobile phones, but they were not interested in controlling them.

The participants could give optional feedback in free form text without any directions. Consistent answers were received which emphasized certain kinds of robot skills or missing capabilities were pointed out. The conversational and interaction skills of AIBO were the most wished, but integrating internet services and AI agents were also top-rated features. The participants preferred the interaction with their robots, but sometimes they wanted to watch their bots wandering around hence the better autonomous behaviors were emphasized in the expectations. After some time, people got bored with the robot software and wished to get new updates to keep their engagement since Sony did not release significant software updates for these expensive robots. Surprisingly, the learning capabilities, memory functions, face and object recognition were ranked less important than the top items. Although humans perceive intelligence by these abilities, but these results suggested, a social robot's utility depends on the built emotional attachment by the interaction skills.

Discussion

After the quantitative and free-text analysis, social robot design recommendations were proposed to improve the market potentials of future products. Although the literature emphasized the importance of a clear purpose for a social robot (M. M. A. de Graaf et al., 2014) to improve the acceptance, but the author argued that a clear purpose is not enough. The customers will abandon their robots if their utility value is under the same value of competing devices in our life, similar to (M. M. de Graaf et al., 2016).

First of all, the robot's appearance must reflect to its capabilities to avoid the uncanny valley (Mori, 1970). Sony AIBO robots are good examples because their resemblance to an animal induced the expectations for an animal-level intelligence. The interactive skills and the conversations were important for the owners to establish an emotional attachment to the robot. The world changed to an Internet-connected society after the emergence of smartphones and the wide availability of data connections in the 2010s. The interest shifted to include web service and conversational agent integrations into the new robots. The owners got bored soon with the repetitive behaviors. Regular content updates are needed for social robots, and nowadays, the mobile app stores and the in-app purchases are successful in the mobile space. Thus, people are prepared to pay for new content. The traditional business models can be extended with the content purchase or monthly subscriptions to ensure future commercial success for social robots. Fortunately, learning and memory skills had a lower priority for the Sony AIBO owners. It is good news for the robotics companies since these problems are among the hardest to crack in the artificial intelligence field. The heavy users of Sony AIBO ranged from teenagers to pensioners, therefore, the target group can include all ages for a new robotics product.

The results of this experiment given a good indication for the expectations of long-term users of social robots, but the participant count posed a limitation for generalization. The participants were invited on a public internet forum and Facebook thus this sampling was not representative for the general public.

3.3.1 Questionnaire Results from Collaborative Development Point of View

The target group of collaborative development was the AiboLife forum members (http://www.aibo-life.org/forums). The author joined this forum for years and participated in the discussions. This online presence gave preliminary credibility among the members before the collaborative development was started as a public project. It is essential to understand the community around an OSS project to ensure long-term sustainability (Barcomb et al., 2018) thus the questionnaire run by the author was important to get know the forum community. The questionnaire participants were from two sources (AiboLife, Facebook), but the majority was forum members and Facebook was only utilized to recruit Japanese participants caused by the language barrier. Figure 12 is a revised version of Figure 11 from Publication II where the answers of Japanese participants (non-forum members) are removed. Figure 12 shows no major differences to Figure 11 except the location, therefore, similar conclusions can be drawn for the community. All responders stated they had good English language skills, which were crucial, like a common language for easy communication during collaborative development. The author assumed before the survey that a typical heavy user is a young male from a Western country, he has a technical profession and he is an early adopter of the latest technologies.

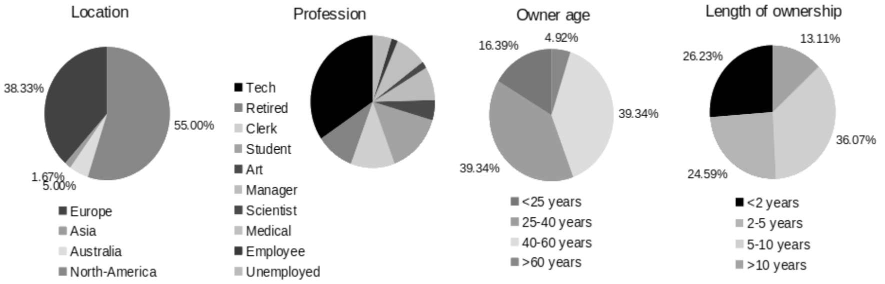


Figure 12. Location, profession, age distribution and length of ownership among the Sony AIBO users at AiboLife forums.

The first assumption was correct, 98 % of the community members were based in Europe, North-America or Australia and this finding reflected to the retail markets where AIBO robots were sold by Sony. It was a bit surprising that 30 % of the responders were female (18 out of 60) because the author expected a lower female proportion. It can be explained since the Sony AIBO robots were designed to build emotional bonds with their owners and women are more capable to build emotional connections than men. This social robot design goal is aligned with these gender attributes and the female owners counted these robots as companions (H1 hypothesis in Publication I).

Forasmuch as the women are less enthusiastic about technological topics, the author assumed, they were less likely to contribute to the technical parts of the AI development. The next assumption regarding the age was not correct as well, 44 % of the responders were more than 40 years old. The older people in the community were not considered as a primary target for technical contributions (similar to women) by the author. These expectations were examined by a question to "contribute new AI behaviors by motion editing" and it was verified for the women. 27 % of them gave a positive or a tentative answer to this question compared to 63 % of the men. However, 55 % of the older people gave positive answers compared to 50 % of the young people. This result shows that technical contributions to an AI are more likely from men than women, but it can be expected by a similar chance from all people considering their age. This latter was also supported by the disproven H2 hypothesis in Publication I.

The reported professions showed a large diversity, but the biggest group was technical (34 %) and the tech jobs were overrepresented among the men (46 %) as it can be expected. This largest proportion was beneficial for likely technical contributions though the majority of the heavy users were employed in non-technical positions. 62 % of the peripheral contributions were code-related at Github (see Chapter 2.1) and 84 % were in the OSS projects of Aldebaran Robotics (see Chapter 2.3). The technical contributions were increased by 22 % for specialized robotics projects over the average OSS statistics of Github. However, Github experiences shown that non-technical contributions can be expected in a variety of tasks.

The length of ownership diagram shows that the community consists of a wide range of heavy users. Some owned their robot for more than 10 years, some owned for less than 2 years. This result was surprising because Sony withdrew AIBO robots from the market for many years when the AiBO+ project was started. This phenomenon can be explained that older members kept their robots after the discontinuation, however, new members joined the community who acquired their robots on the second-hand market. The continuous, healthy renewal of the forum community ensured that there were always newcomers inside the community who might be possible contributors to AiBO+. The rejection of the H4 hypothesis in Publication I demonstrated that the heavy users did not lose interest in their robots after many years. Therefore, the contributions can be expected from the whole community, regardless of how long they have been owning their robot. 59 % of the owners considered making new content for the robot. Mostly male community members (27) expressed positive answers compared to 5 females and 43.75 % of these responders had tech jobs. These results indicate that males have an overwhelming 5x chance for contribution and the members with tech jobs were overrepresented compared to their proportion in the community (43.75 % vs. 34 %). Surprisingly, the older the heavy user was the higher the probability of contribution was and the longer ownership also increased the chance.

On one hand, the forum community is a target to engage members for contribution, on the other hand, they are also the consumers of the developed features. The questionnaire answers gave some clues about which AI features the robot owners would like to encounter in new software.

As the analysis in Publication II pointed out, the community members were not interested in the repeated, non-interactive entertainment features like dances. Because of this reason, producing new motion behaviors was not a good time investment for the author to engage people. They appreciated the human-robot interaction skills like speech recognition or distinguishing humans. The robot's camera had low resolution (0.3 MP), poor quality for average environmental illumination, and rolling shutter. Therefore, the audio-based improvements were implemented by sound event recognition (Section 3.4.2). Autonomous features were preferred and they were a natural goal for the author because these AI skills give the impression of the intelligence for humans. Additionally, the owners wished to connect their handsets to their robots. It was understandable since AIBO robots were discontinued in 2006, right before smartphones became widespread after the iPhone's success. The author implemented an application for mobile phones to reach the dog via a wireless network (see Section 3.4 and 3.6). The continuous development of AiBO+ was important, and the author devoted years to it, as the community members missed the regular software updates for their robots from Sony.

The Sony AIBO product line had several models over the years. The SDE was universal to develop C++ programs for these robots, but they had physical differences and the various sensor configurations. It was not practical to develop the same software for more models, especially, since the AiBO+ project was started by the author alone. The ERS-7 robot was chosen as a target platform because this model had the best hardware and it was the most loved by the community (see Publication I).

The programming language for AiBO+ was determined by the shipped SDE from Sony and the C++ language was the only choice. This programming language is well-known for the ability to write low-level, high-performing programs for embedded platforms. However, it has a long initial learning curve and high language complexity. These obstacles limited the project involvement for code contributions. Although 43.75 % of the community had tech-related jobs, it did not imply these people were software developers and proficient in C++. However, familiarity with computers can still facilitate non-code contributions by resolving technological problems more easily.

3.4 Artificial Intelligence Features For User Engagement

An AI must provide at least an impression of an intelligent agent and it was discussed in Section 2.1 that the robots must avoid the uncanny valley by maintaining reasonable user expectations. This design choice can be observed in the Anki robots (see Section 2.1.1) because they created a silly, childish personality for their toy robot with cute aesthetics. Since the Sony

AIBO robots resembled a dog, it was easier to satisfy the expectations for animal intelligence. These robots were expected to execute autonomous locomotion since they have four legs. This feature is necessary to avoid such a situation when the participants were disappointed by Pleo robot in an experiment because the robot had legs, but it could not walk (Fernaeus et al., 2010). The AIBO robots do not have a real utility purpose, they just act as a companion to humans. These robots can explore the room on its own, dance and play funny motions to entertain people what requires environmental sensing. The situation awareness and human recognition are necessary for human-robot interactions. The initial content was implemented by a few AI features (Table 1) to provide a basis for engagement.

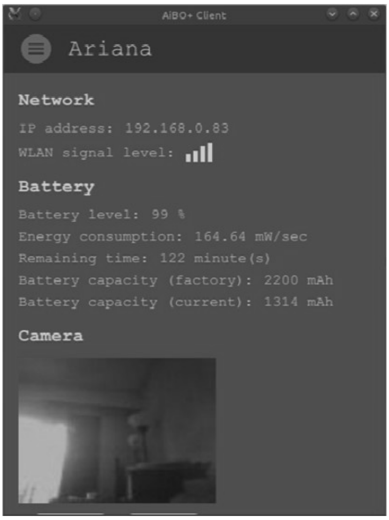
AI Feature	**Details**
Locomotion	The walking and turning motion sequences were borrowed from an original Open-R programming example (MoNet). Motion transitions between basic postures came from the motion library in Skitter (see Section 3.5) created by the AIBO community. Recovering from fallen down, avoiding abyss and obstacles were implemented in Section 3.4.1 and Section 3.4.3.
Human-robot interaction	A number-guessing game can be played with AIBO via an Android or iPhone mobile application (*AiBO+ Client*, 2020). The robot begs to place on the ground from being picked up and it asks to be placed on the floor from a sofa or a chair based on the embodiment awareness (see Section 3.4.1 and Section 3.4.3). Emotions are expressed with LEDs and audio effects according to the Sony specifications to avoid the confusion of the owners and to experience familiar reactions from the robot.
Perception and reaction	Sound event recognition was implemented (see Section 3.4.2) to explore interesting events in the surroundings. The underlying surface was detected with a new method (see Section 3.4.3) and this knowledge was used to change the walking speed on different surface materials (e.g. carpet, vinyl). Loud sounds could be discovered by turning the head towards them to get a camera image of the source. The LED brightness

	(emotion and state indicators) was adjusted according to the environmental brightness and the volume levels were adjusted according to the time of day.

Table 1. Artificial intelligence features for the initial content.

The engagement for collaborative development is not easy. The users must initially be excited to get a wow effect and look after the project. After this stage fades away, the engagement must be kept up over time. Since this study works with heavy users of Sony AIBO who are in their robot acceptance phase, their engagement requires less effort than a random user population. The author's technical offerings (embodiment awareness, audio and environmental sensing) were provided enough new innovative ideas that were worth checking. As a result, the user base of AiBO+ was grown over the years, described in Section 3.5.

An application (*AiBO+ Client*, 2020) was developed (Figure 13) for multiple platforms to connect to the robot and address the owners' wishes that they expressed to the questionnaire in Publication II. The users could view the basic status information of the robot inside the application, play a game, move the robot around and participate in a crowdsourcing campaign for scientific data.



Network

IP address: 192.168.0.83
WLAN signal level:

Battery

Battery level: 99 %
Energy consumption: 164.64 mW/sec
Remaining time: 122 minute(s)
Battery capacity (factory): 2200 mAh
Battery capacity (current): 1314 mAh

Camera

Figure 13. Client application to connect to the robot.

Offering Quality

The software quality of the AI was an important target to achieve a good first impression in the community members. Since the complex robots are expensive, a faulty AI is not acceptable because it can cause personal harm or property damage. It was shown in (Garza, 2018) that the failure severity and frequency are influential factors for robot acceptance. While most participants assisted the robot when it failed, they were more likely to help when they had not observed other failures earlier. This phenomenon suggested that the users are tolerant of failures, but the possibility must be minimized for any wrong behavior.

As the HBS study of Jibo (Bussgang & Snively, 2015) shown in Section 2.1.1, the fulfilled promises are important for social robots. Therefore, before AiBO+ was publicly released for other robot owners, the author spent several years building the software basis with the initial offerings.

The software development for Sony ERS-7 is constrained. A small amount of memory (64 MB) and computational power (600 MFLOPS) are embedded inside this robot. These specifications can be compared to a modern smartphone which reaches over 100 GFLOPS and it has gigabytes of RAM. The main processor processed all operations of the robot (video, audio, motor control), therefore, all software functions must be run in real-time to spare with scarce resources. Since the memory is small to run a program and store data, the memory management inside the AI must be perfect to avoid any resource leaks. The SDE for Sony AIBO robots was released under Linux for a host computer and the supported programming language was C++. Linux has a wide variety of tools to ensure the correctness of an application, but the Aperios system lacks any tool to diagnose the problems except a crash tracer which was rewritten by the author in (Kertész, 2013). The best available programs and methodologies under Linux were utilized to write a bug-free AI. Valgrind and Clang sanitizers tracked down all memory handling problems, gdb was useful to debug software crash problems. As it was mentioned for Aldebaran Robotics in Section 2.3, the AI must be tested without the robot to protect the expensive hardware from damages. To reach this goal, architectural, modular (Garcia et al., 2005), singleton (Stencel & Węgrzynowicz, 2008) and observer design (Ghaleb et al., 2015) patterns were implemented to define functionally separated AI modules. The AI behaviors, the actuator controllers were developed in distinct components and integrated with a glue layer to the robot. Except for this latter part, all other codes could be compiled on the host machine and tests were defined to verify the correct results with mocked hardware functions. With the aforementioned techniques, the low-level functions were guaranteed to work properly, but the higher-level AI features required other verification to ensure the proper functionalities.

The reproducibility is an important part of testing AIs because the growing complexity implicates the increased chance of failures. The behavior-based robotics is a common design principle in research and industry. It decomposes the problem solving into small behaviors where each solves a subproblem inside the problem domain. Although this approach helps to handle the complexity in AI, it still demands continuous testing to verify the overall correctness. The author defined the deterministic test cases (DTCs) by extending the standard software testing practices for robotics and DTCs could test incremental changes in the new AI for Sony ERS-7 (Kertesz & Turunen, 2017). Recorded sensor data for every test case and design patterns guaranteed that the behavior execution was equivalent on the robot and its simulated counterpart on a laptop. These DTCs were created easily, their execution was fast and they minimized the manual testing with the physical robot.

An AI project can be made more reliable after choosing an appropriate open-source license and publishing it on the internet. In this way, the project gets feedback, reports about problems and contributions. AiBO+ was published on a webpage for OSS (http://aiboplus.sourceforge.net) and the source code was shared under GPL 2 license.

A Short Introduction to Machine Learning

The next three articles have different problems, but the same machine learning methods were used to solve them. In machine learning, computer programs are taught to learn a suboptimal solution by creating a model on sample data called training data (Russell & Norvig, 2009). Each training sample has a certain label (e.g. cat or dog for pet images) and the model learns to assign the correct label for the training samples. Once the model is built, the computer can make predictions inside the problem domain by looking at new, unseen data (testing set). For example, a cat-dog classifier model can decide if a new picture contains a cat or a dog (Huh et al., 2016). The more samples are classified correctly the better the model, however, perfect model does not exist, there will be always some misclassifications (Sünderhauf et al., 2018). The models are ranked how accurate they are by the accuracy metric that is the correct prediction count divided by the overall prediction count. 100 % accuracy is a hypothetical perfect model when all predictions are correct, but the accuracy in the 85-99 % range is considered satisfactory.

The problem-solving with machine learning is generic, it can be applied to different kinds of data. Section 3.4.1 deals with robot body state recognition based on sensor data, Section 3.4.2 implements sound event detection and models are built on sensor data to detect the underlying surface in Section 3.4.3. In all cases, the models are transferred to the robot to make predictions in real-time.

3.4.1 Self-Awareness

The embodiment and self-awareness are important for robots. The consumers are more accepting of these machines if they feel empathy for the embodied agent because of its aesthetics and behaviors. User acceptance is higher when self-awareness is reflected in artificial intelligence, therefore, it is essential to implement such features in the content offerings. The author implemented a basic body state recognition based on the embedded accelerometer to identify situations when an anomaly happens. The normal circumstances included lying, sitting, standing poses and walking. The work in the following article describes how the robot detected when it was picked up by a human to carry around or when somebody undesirably pokes the robot from the side as well as the robot recognizes after it was fallen down during locomotion. The robot either reacted to the latter situation by getting up and walking further or asking the human to stop disturbing him in the other situations. The robot also gave visual feedback with the onboard LEDs when the unusual body states were detected according to the Sony specifications to show familiar feedback from the robot.

Reference

Kertész, C., & Turunen, M. (2018). Body State Recognition for a Quadruped Mobile Robot. *Proceedings of the IEEE 22nd International Conference on Intelligent Engineering Systems*, 323-328.

Objective and Method

The legged robots are normally in a stationary position or walk around. They can make appropriate reactions to irregular events if they recognize their own body states, therefore, three anomalies were taught to a Sony AIBO to recognize when it is picked up, fallen over or being poked. Since these events involve specific patterns of the body movements, the accelerometer was utilized in the robot.

A dataset was collected by the accelerometer while the robot operated. To record the *normal state,* the robot walked around, lay down, sit and stand. The *fall over* events were recorded when the robot lost the balance and fallen on a pillow to avoid any damage. Two human-related interventions were considered as anomalies. Once a human picked up the robot and walked around with it, the robot became in *picked up* state. When the robot was hit by a human from side, it is entered in *poked* state. 76535 samples were in the dataset, 52612 samples for normal, 12831 for picked up, 1323 for fall over and 9769 for the poked state. The normal state had most samples because it includes a variety of activities (static poses, walk, object manipulation etc.). Poked and picked up states were still easy to record, but the fall over samples were the least since it was dangerous to the robot and one fall over event generated just a few samples. The dataset was separated randomly to training and validation sets (58%/42%). The machine learning models were built with the training set and the 10-fold cross-validations were run on this set while the validation set contained the unseen samples shown to the model.

A low-end accelerometer in the robot generated the feature vectors for the machine learning methods. A 270 msec-long window was sliden over the sensor values and statistics (interquartile range, min, max) were calculated for each accelerometer axis. Above these 9 statistics, 4 values described the robot intentions for locomotion. Six classifiers were examined with 10-fold cross-validation and model evaluation to find the optimal solution with satisfactory accuracy and real-time requirements.

Results

Support vector machine (SVM), naïve Bayes, k-nearest neighbor (KNN), decision tree, random forest and deep neural networks (DNN) were explored to find the best machine learning model for the body state recognition. When the training set was run with 10-fold cross-validation, the SVM and naïve Bayes performed under 80 % accuracy, DNN close to 90 % while KNN and decision tree-based methods were close to 100 %. Model evaluation on the validation set changed the order. Deep learning was the best method with 98.01 % accuracy followed by SVM and KNN. Decision tree and random forest achieved moderate results while naïve Bayes was under 50 %. For the deployment in the robot, accuracy is important, but the selected method must be executed in real-time. When the best-performing algorithms were compared, SVM had the lowest memory consumption (2 KB) and execution time (2 usec). Although the deep learning model needed around 10 times more resources, but 28 KB memory usage and 20 usec inference time were still very low paired with the best accuracy on unseen data. Therefore, this model was selected for the robot and runtime evaluation confirmed that it performs in-the-wild correctly.

Discussion

Several classifiers were evaluated for the described problem and the deep learning model delivered the best accuracy and satisfactory runtime performance. A closer look at the confusion matrix of the DNN model revealed that almost all events were recognized well, but 13 % of the fall over events were classified as normal locomotion. This case can be handled with late-stage filtering since none of the other events were misclassified as fall over thus if an input vector was classified as fall over it was likely true. All other states (normal, picked up, poked) were recognized with 94-99 % accuracy. Although temporal deep learning models can enhance the results further, but the current model was composed of two simple fully-connected layers with 20 neurons and this shallow network could deliver practical performance shippable to other robot owners.

3.4.2 Audio Sensing Modality

Although the sound is an important sensing modality, the research teams usually take off-the-shelf solutions to recognize human speech to give conversational skills to their robots. Since the Sony AIBOs resembled a dog, they did not have to excel in conversations with humans. It is enough to recognize some simple voice commands and detect certain events in the environment. The official Sony software provided 100 voice commands to interact with ERS-7, but they did not include any environmental understanding by audio cues. Providing this missing feature was a good incentive to develop to engage the community members for contribution. The author considered some common sound events in the living rooms where these robots were intended to work. Humans, animals, instruments and household appliances were the target sounds to distinguish from the ambient noises which were also modeled. When these sounds are recognized, they are audio sensing blocks in the machine intelligence to execute specific behaviors or interactions.

Reference

Kertész, C., & Turunen, M. (2018). Common Sounds in Bedrooms (CSIBE) Corpora for Sound Event Recognition of Domestic Robots. *Intelligent Service Robotics*, 11(4)335–346. doi:10.1007/s11370-018-0258-9

Objective and Method

The robots have limited computational resources and power, but they must respond to the environmental stimuli in reasonable reaction time. Real-time sound event recognition by deep learning was explored in this article for Sony ERS-7 robots. 13 sound events were chosen to be recognized while omitting the background noises. The whole task was approached with balancing between onboard processing and good detection accuracy.

Since there was no suitable audio corpus available for indoor applications, a new dataset was collected from free online databases, public research datasets and some new recordings were made. The collection of these original raw sounds was the first part of the database and it was named to CSIBE-RAW.

The application on the robot requires a dataset that incorporates the microphone dynamics, various noise levels and reverberant conditions. To solve this problem, CSIBE-RAW was rerecorded in four settings. The raw samples were played back by a high-quality speaker and the robot's microphone recorded it. Two settings were taken place in a reverberant room where the speaker was 1 meter away 30° counterclockwise to the robot head and the other setting was 3 meters away, 1 meter high and 180° clockwise to the head. The other two settings were the same except they were in a non-reverberant room. This part of the database was named to CSIBE-AIBO.

Features must be extracted from the audio to apply machine learning to this problem. The audio data were framed by a sliding Hann-filtered window (32 msecs) with 33% of overlap, fast Fourier analysis was performed to extract 23 scalar statistics and 26 MFCCs for each frame. The first MFCC coefficient was dropped, the remaining were added to the feature vector which contained 48 features in overall. These features were quick-to-compute (1 msec) on the robot and they were robust to lossy audio codecs. The sound event detection was frame-based (32 msecs), but the predictions were temporally smoothed by majority voting.

Results

The CSIBE-RAW and one setting from CSIBE-AIBO datasets were randomly separated into a training and a validation set. The training set was used for cross-validation while the validation set acted like unseen data in the later evaluation. Nine classifiers were examined for the problem: maximum entropy, support vector machines (SVMs) with linear and radial basis function (RBF) kernels, naïve Bayes, k-nearest neighbors, decision tree, random forest, expectation-maximization and convolutional neural network. The classifiers were checked by 10-fold cross-validation. SVM with RBF kernel and expectation-maximization classifiers had low performance hence they were left out from the forthcoming evaluation. All other classifier reached accuracies over 80 %. Then a feature vector transformation method was implemented to improve the generalization power of the models. The feature vectors were replaced with aggregated frames which were computed by the mean and standard deviation of every 9 subsequent feature vectors with a 30% overlap. This data aggregation resulted in smaller training set size and improved accuracy compared to the frame-based evaluation. This baseline system was tested on other datasets from the literature where they reported their own cross-validation results and the author method outperformed the results in the literature.

The cross-validation estimates the real model accuracy with unseen data, but it can lead to misunderstandings about the generalization power of a classifier. Therefore, seven classifiers were evaluated on the validation set and all classifiers reached good accuracies with 90-96 % after majority voting. The best performer was the CNN classifier. The cross-validation underestimated the actual model performances because almost none of the cross-validations of the same classifiers reached 90%.

The previous experiments included only one setting from CSIBE-AIBO. The next challenge for the classifiers was to handle all four settings in this dataset. Multi-conditional learning was selected to build a robust model for different SNR levels and reverberant conditions in this article instead of generating synthetic training data. The CNN classifier had the highest accuracy again with 95.07 % and it run on the robot with 3.5 MB memory usage and 6 msec execution time.

Discussion

The audio cue is an important modality for the robots to sense the environment and give appropriate reactions. Since the robotics field lacked a domain-specific indoor dataset, the author created the CSIBE corpora for non-overlapping sound event recognition. The database contained 14 sound event classes where 13 events covered human speech, animal voices, musical instruments and household appliances. One additional class modeled the ambient noises (e.g knock, drawer, keyboard, paper, breathing, steps) which are not important for a domestic robot. The various reverberant conditions and SNR levels pose a challenge for sound event recognition. The author solved this problem with multi-conditional learning although there are other synthetic data augmentation techniques to handle these situations. The original sounds were rerecorded with the stereo microphone of a Sony AIBO robot in four room settings and a deep learning model (CNN) could reach 95.07% accuracy with unseen data and the model could be deployed on the robot with real-time capabilities.

3.4.3 Environmental Sensing

The major focus of the official Sony software for AIBO robots was the entertainment aspect aligned with product marketing. The human-robot interactions, playing with toys, dancing and tricks were the main content. Although the ERS-7 robot could avoid obstacles and go back to the charging station if it was in the visible range, the onboard software did not have any advanced environmental mapping or context-awareness. The author implemented some behaviors with utilizing the infrared sensors to enhance the responses in certain scenarios:

- The robot refused to leave lying pose if abyss was detected before the robot.
- The robot turned and walked away if an abyss was detected while walking.
- The robot detected nearby obstacles before colliding into them, it turned and walked away.

The fourth environmental sensing feature was detecting the underlying surface with machine learning and sensor fusion while walking. The details are described in the following article.

Reference

Kertész, C., (2016). Rigidity-Based Surface Recognition for a Domestic Legged Robot. *IEEE Robotics and Automation Letters Journal,* 1(1)309-315. doi:10.1109/LRA.2016.2519949

Objective and Method

Surface rigidity is an important knowledge that can be applied to switch to a more suitable walk pattern or walk speed. A Sony ERS-7 robot walked with singular crawling and a 2400 msec-long walk period. The samples were collected on tiles, wood flooring, vinyl flooring, carpeted floors, short carpets and soft carpets. These six surface types are listed here in order from hard to soft and multiple examples per type (e.g. more carpets) were shown to the robot to achieve better intraclass variability. The body oscillations were used to build machine learning models to detect the different surfaces and the robot worn socks or walked barefoot during the sample collection. The original intention was to develop two separate models for socks usage and barefoot, but the initial experimentation showed no difference in accuracy when these samples were separated or used together to build models, therefore, one model was trained with collapsed samples.

Infrared, motor force, ground contact force sensors and accelerometer were fused to distinguish six domestic surfaces based on rigidity in this article. The accelerometer in the Sony ERS-7 is a low-cost model with a 120 Hz sampling rate. The infrared sensor had a 25 Hz sampling rate and was directed to the ground by 30 degrees. The ground contact force sensors were simple two-state buttons in the paw with a 10 Hz sampling rate. The ERS-7 has force sensors in each leg joint to measure the mechanical load in the joints. The hip joints of the hind legs were found the largest discriminative power out of the other joints.

Two walk periods (4.8 seconds) of sensor data were the basis to generate feature vectors in a sliding window to catch the relevant body oscillations on a certain ground surface. 30709 samples were collected in the dataset and the author published on the internet for free. The corpus was split into a training set and a validation set randomly in 40 %/60 % partitions. The classifier performance was estimated with cross-validation on the training set and the final models were also built with this set to be evaluated by the validation set as unseen data.

As it was mentioned before, the feature vector was built on the sliding window of the sensor data. Fast Fourier transform (FFT) was computed over the sensor data of different modalities. The most useful frequency bands were in relation to the walk period of the legged robot. Namely, its overtones and the inharmonic partials hold the most information for surface classification, confirmed by a feature selection method. Six FFT components were used for the accelerometer axes, five for the infrared sensor and three for the force sensors. Some other scalar statistics (median, maximum, skewness, interquartile range, sum etc.) were additionally computed over the sliding window of the sensor data and the results were added to the feature vector.

Six classifiers were evaluated to find a suitable method with good accuracy and quick execution on the robot: maximum entropy (ME), support vector machines (SVM), k-nearest neighbors (KNN), decision tree (DT), random forest (RF) and kernel ridge regression. The classifiers were evaluated by 10-fold cross-validation on the training set to estimate the accuracies on unseen data. Once the results were available, models were built on the whole training set and they were evaluated on the validation set as unseen data.

Results

ME, SVM, and KNN had accuracy over 80 % in the cross-validation, but DT and RF reached the best accuracy with 91-96 %. The evaluation on unseen data shown varying results, the accuracies decreased for most classifiers as expected. The random forest had the most stable performance across the use cases, it had the best score among the other algorithms and the validation accuracy (92.19 %) was close to the cross-validation performance (96.29 %). Due to these results, RF was chosen for the final experiments to get an optimal model with a balance on accuracy and real-time speed.

Depending on the hyperparameters of a random forest, it can underfit the data if the forest is too small or overfitting happens if the size is too large. The author arranged a hyperparameter search to explore the accuracy and the memory usage to optimize these parameters for the onboard execution on the robot. The experiments showed an accuracy plateau of 91-94 % for higher parameter values, but choosing the minimal values was the target to have low computational complexity. Finally, both parameters were set to 20 and the final accuracy was 91.57 %. The confusion matrix of this model revealed that the most notable misclassifications happened between surfaces with similar rigidity.

An inherent feature of the random forest is that it can provide feature importances to determinate which features contributed the most to the predictions. The weak predictors could be removed from the feature vector and the removal did not affect the model accuracy negatively. Every modality was significant on average, but the accelerometer's z-axis produced relatively weak discrimination power that was against the experiences in the literature. Possible causes can be the different extracted statistics from the accelerometer data or non-equivalent mapping of the three axis data between the paper and the literature. Future investigations might be helpful to reveal the real reasons more accurately. The maximum and 3rd statistical momentum over multiple sensors provided stable, average discrimination. The RMS amplitude statistics had good performance among the other features and the ground force sensors were outstanding despite the fact that they are simple two-state sensors.

The final computational requirements were low. The training was executed in a few seconds on an average laptop, the feature extraction took 3 msec on the robot's embedded processor and the prediction time was 20-90 µsec paired with 833KB memory usage.

Discussion

A process was described in the article to design a surface recognition model. All available onboard sensors (infrared, motor force, ground contact sensors, accelerometer) were utilized to distinguish six surface types with high intraclass variability and walking on them barefoot or in socks. Several classifiers were evaluated with cross-validation and the random forest was selected because of its stable performance. The minimal model size was determinated by the hyperparameter search to get a balance between performance and computational complexity. The feature importances found the most useful features to build a model without losing accuracy. The final model was deployed to the robot with low computational requirements.

The experimental experiences and the literature review revealed a few recommendations for future researches in this area. The FFT coefficients have good discriminative power for sensors of different sensing modalities. The use of FFT components with the overtones and the inharmonic partials of the walk period are advised. Even the simple ground force sensors predict the surface rigidity very well. Root mean square amplitude, interquartile range, median, skewness and maximum are recommended statistics for feature extraction in this research domain. The random forests can be used for the preliminary experiments and feature selection to get experience with a particular robot and collected dataset, but powerful features will provide similar prediction power for different classifiers.

As the short review of this article shown, the author developed with environmental sensing skills successfully for the robot and it could become part of the content offerings.

3.5 COLLABORATIVE DEVELOPMENT RESULTS

The author aimed to maximize the community contributions, therefore, an open and transparent development process was established. AiBO+ was registered on a hosting service for free software (http://aiboplus.sourceforge.net) because people contribute to open-source software with a higher chance than a project with a restrictive license (Crowston et al., 2012). Not only source code was released under this open-source license, but other content contributions (detailed in 3.5.2) as well. Since the author participated in the forum discussions, he made the active development for the source code and these actions gained trust in the forum community. The author did not analyze the community as an external researcher, but he played an active, internal role in the whole experiment. Hence, this dissertation can be distinguished from other earlier works in which the researchers examined online community activities from an outsider perspective (Fink et al., 2012) (Kahn et al., 2002).

When the first AIBO robots were released around 2000, the internet was not so widespread and the free software was a relatively new phenomenon in the software industry. As these foundations of the collaborative development were missing, the online communities were not established at that time. 3rd party programmers were not involved in the corporate innovation and the companies did not plan direct communication with the customers via online channels. The lack of these internet resources encouraged the Sony AIBO customers to build these web pages by themselves. Several homepages were built with forums (http://www.aibo-life.org, http://www.aibosite.com, http://aiboworld.tv). They used these sites to share their experiences, discuss topics about their robots and collaborate in certain projects. This self-organization continued after the AIBO robots were discontinued and repair services were initiated (http://www.homerobot-service.com, http://www.aibohospital.com). The AIBO community built their basic online services without corporate support because these strong robotics products were ahead of their time and extremely engaged their customers.

The author started AiBO+ and bought his robot dogs years after the Sony AIBO robots were pulled from the commercial market. Although the project had to be built up, it did not mean to create everything from scratch, contents could be reused from two sources. On one part, the limited software updates to the robot dogs were present since the early years of the 2000s and the robot owners felt the need for new content. Obviously, Sony did not provide it, the community created new things which were technically feasible. A community member (DogsBody) wrote a program called Skitter that could edit motor motions and LED animations. Skitter could test these sequences on a simulated 3D model of the robot and send the whole skit to a physical robot for playback. Once this application was built, the community members started to create different dances and funny motions for the robots. The author used this software for motion editing extensively during the dissertation project. The second main source came from Sony after they opened their SDE to the public. Although they did not let the users adding new functions to the existing AIBO software, but they released sample applications to show the potential in the software development for their robots. One of their examples provided straight and turning walk patterns. They were extremely important since developing walk sequences for quadruped robots is time-consuming, threatening the expensive robot hardware and it is a research topic on its own. The author embedded these walk sequences into the initial offerings of the AI engine.

The author was the core developer of AiBO+ started in 2009 before the collaborative development was established with the community members on the AIBO forum in 2015. All other contributors from the AIBO community were peripheral developers from the project point of view. Their contributions are detailed in the upcoming sections.

3.5.1 Code Contributions

The code contributions were unlikely to AiBO+ due to the previously mentioned software development difficulties for AIBO. It took 5-6 years for the author to implement a reliable basis for the software development, simple AI features for the initial content offerings and client applications for Android, iOS, macOS, Windows and Linux to connect to the robot via a wireless network. 91432 lines of well-commented source codes were implemented over 9 years, 92 % of them were written in C++. These efforts equal to 23 man-years and about $1.2 million costs. Detailed statistics are available at https://www.openhub.net/p/aiboplus.

This new AI software had to work together with other existing software to ease collaborative development with other members. Skitter was a user-friendly application for motion editing that was written before 2010 and it communicated with the robots via TCP ports on the wireless network. AiBO+ used a more reliable UDP network communication for the same purpose thus they were incompatible with each other. The author contacted the developer of Skitter (DogsBody) to add support for UDP communication and switch to this mode on-fly when UDP broadcast messages are detected from the robot on the network. He kindly agreed with the contribution and he implemented this new feature into his program and fixed a few bugs found by the author. Since the Skitter program was not a paid application, this application was bundled inside the downloadable distribution packages of AiBO+. Because these code implementations were a one-time contribution to AiBO+, DogsBody was an episodic developer.

3.5.2 Content Contributions

The content contributions were started before the AiBO+ project. The author of the Skitter program collected all community-created motions and included inside his application. Finally, 281 new motions and skits were constructed by multiple community members over the years. These community-created motions were valuable because motion transitions were available between basic poses (e.g. sit, stand). Since the motions were created with different AIBO models and the mass distributions vary between them, these community skits could be used for ERS-7 after some editing.

The author was uncertain what kind of contributions the non-technical members can make, but the community itself suggested their contributions to the AI. Namely, a few members proposed to voice act for AiBO+ because the robot used a text-to-speech engine to speak to the humans and it sounded bad. In the end, two female members voice acted over several years. The process was simple for these contributions. Before the author made a software release, the new English sentences were sent to the voice actors, they recorded the content on their computers and the records were sent back in an email. After some minimal editing, the sounds were converted into an internal data format and included in the new AI software release. The voice actors were engaged with AiBO+ for years, but new software versions were released once or twice a year, their contributions were episodic.

3.5.3 Other Contributions

As mentioned before, the AIBO community invented a solution spontaneously when they came across any problem. The discontinuation of the AIBO products led to the end of all warranty periods from Sony in 2013-2014. The robot owners were left without a source of spare parts and official repair services, so a solution was needed. In the 2010s, 3D printing emerged into the mainstream hobby space and 3D printing of plastic pieces became affordable with a prototyping process. These services are available at universities and libraries nowadays. Besides the electronics of the Sony AIBOs, many plastic parts are embedded into these robots. The author collaborated with other community members to build free 3D printing models for replacements of missing parts and such models are available on Thingiverse (https://www.thingiverse.com/search?q=aibo): ear, ear clip base, station pole and pole base for Sony ERS-210, battery shell for Sony ERS-11x, knee joint for Sony ERS-31x, Aibone (toy) and station pole for Sony ERS-7 (Figure 14). Each creator built one 3D model, therefore, they were episodic contributors.

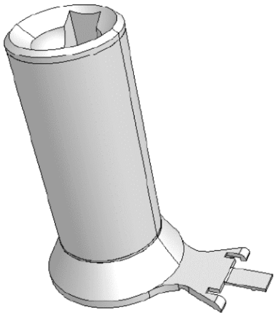


Figure 14. A 3D printing model of the Sony ERS-7 station pole.

Crowdsourcing is a method to obtain services, ideas or content by collecting contributions from a community. The author applied it to this research as a voluntary data contribution to solve machine learning problems. The robot owners downloaded the AiBO+ client application (AiBO+ Client, 2020), connected to their robot, recorded samples for different problems, and shared with the author for scientific analysis. The participants got a notification inside the application when their contributions were uploaded, but there were no reward mechanisms like point systems to drive these actions. This schema was used to gather the dataset for surface recognition in Publication IV.

Another type of non-code contribution is a bug report. When a user encounters an error in the software, he puts effort into entering a summary about the problem for the developers to fix it. There were two major software bugs found and reported by the community members. The AI crashed when the WLAN switch was off on the robot during startup and the macOS could corrupt the directory names when the AI engine was copied to the memory stick. Both problems were fixed in some days and a new version with hotfixes was uploaded to the OSS project site on SourceForge. These quick response times build trust in the author among the community members.

3.6 PROJECT EVALUATION

The author set a goal for success at the beginning of the thesis project to involve multiple AIBO community members in AI development and incorporate their work. As the description of the community contributions showed, this target was achieved and several forum members became peripheral contributors to AiBO+ for more years.

Although one can argue that this project involved a limited number of peripheral contributors, the author was the only core contributor thus we cannot expect a commercial level AI engine which can attract many people. For example, AiBO+ can be compared to Anki's Cozmo robot (Table 2). As

Cozmo was created by a startup company and professional employees, there were many more content and AI features to offer to attract people for contribution. Nevertheless, despite the small AIBO community and the simple AI engine developed for this dissertation, several members still sacrificed their leisure time to contribute to AiBO+.

	Cozmo	**AIBO community**
Motion animations	2000	281
Audio content	42-min original music	Voice acting
Source codebase	2 million lines of code	91000 lines of code
Motion editing	Maya editor	Skitter
Client applications	Yes	Yes

Table 2. Feature comparisons of the commercial Cozmo and AIBO community projects.

AiBO+ popularity can be measured directly if we look at the client application's analytics on the Android Developer Console (Figure 15). The active installations were stable around 40-60 between Sept 2016-Dec 2018, but then there is a continuous growth since Jan 2018 reaching the range 90-100. This latter increase is explained by the release of the newest AIBO product (Sony ERS-1000) to the market on 11 Jan 2018. This new model gave an extra level of awareness for Sony AIBO in the markets and the popularity of AiBO+ was increased since 2018 in spite of the fact that the last version upgrade was released in Jan 2017. The app ratings can also measure the success and 14 users left an average score of 4.6 out of 5 on Google Play as of 2 Oct 2020.

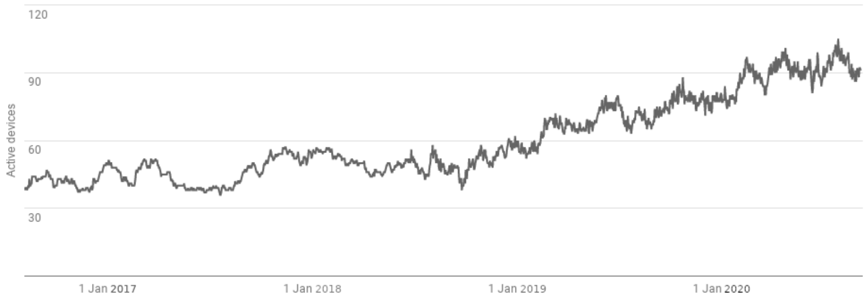


Figure 15. Active installations of the client application on Android devices between 4 Sept 2016-2 Oct 2020. The time period before Sept 2016 is not available on Google Play.

3.7 Discussion

Although the collaborative development in this dissertation is novel to involve people on the Internet for AI development, the online communities were also explored in the past to incorporate them in product innovation and distributed open-source software (OSS) development.

Industrial Community Involvement and Crowdfunding

The primary goal for user involvement in the 2000s was to understand the customer needs better and gather ideas from an external source to reduce the development costs and shorten time to market (Greer & Lei, 2012). When this involvement was more profound, the customers were an active part in developing a new product (da Mota Pedrosa, 2012) and the brand could use it as an advantage in the marketing. Customer and user involvement were studied in many works (Coviello & Joseph, 2012) (Sandmeier, 2009) and the results strengthened the benefits for product innovation. Technological innovation can only be successful when consumer expectations are understood during the process. The user requirements must be collected and tracked over time with active communication to the clients (Hemetsberger & Godula, 2007). Customer involvement can reduce development costs, invent new ideas, and be an active member in the product development process (Fang, 2008) (Greer & Lei, 2012). However, involving customers have effects beyond product development. When some customers commit early purchases and share their experiences on the Internet as reviews, the product will get preliminary credibility on the market.

The business models of Indiegogo and Kickstarter are based on this principle. The early adopters support projects in the prototyping stage on these websites and can receive a product if the crowdfunded project is successful. The robotics companies recognized the power of these platforms as proper mediums to reach tech-savvy, early adopter consumers. However, announcing a project on Kickstarter does not yield an inherent success, the companies must treat these platforms as an integral part of their strategy and product development process. The crowdfunding platforms have a good advertising value to get initial consumer attention, but the project is doomed to fail if these platforms are purely treated as marketing means to attract more consumer money. The early adopter feedback should be injected into the company processes and clear communication can keep up the positive consumer attitude. There is a positive early example from the motorcycle manufacturer Harley-Davidson before the crowdfunding web-based services were founded. The chopper owners discussed their product experiences in an online community and the company included the community ideas into the development processes (McAlexander et al., 2002).

Crowdfunding a project always has challenges. Skovgaard and Gurzan (Jensen & Özkil, 2018) studied 114 projects on Kickstarter to identify engineering design and product development issues in these campaigns. The analyzed projects promised consumer products with a funding goal above $5000 in the technology category. In their results, 43 % of the projects were delayed over 12 months, 25 % lacked expected features and 64 % faced manufacturing quality issues. These threats to a successful campaign must be taken seriously since these issues can harm the positive response on the retail market (Hendricks & Singhal, 1997). The robotics companies attempted to include the crowdfunding platforms in their product launch (see Section 2.1.1 and 2.3) without success. The most notable failure from the robotics industry, the case of Jibo. Although they launched a robot by an Indiegogo campaign, their primary intention was only to estimate the consumer sales for the venture capital investors. Their campaign was not included in their original product roadmap, therefore, they could not incorporate the crowdfunding project resources into their daily company operations organically. They did not listen to the early adopters and their overpromising marketing materials did not help to avoid their failure either. Although the author did not launch a crowdfunding campaign, he tried to avoid similar mistakes and executed a questionnaire evaluated in Publication I and II. The results gave input from the community to set the directions in the artificial intelligence development and the owners' feedback was tracked on an online forum to communicate and fix any discovered issues.

Collaborative Software Development

Software development is another way to involve online communities in collaborative value creation. The supporting technologies and services were enhanced for the open-source software (OSS) projects in the past decade to facilitate the free software. Git became dominant for version control and the markdown language democratized the documentation writing for non-technical people. Several homepages provided hassle-free project management with git hosting, wiki pages, bug tracking and code reviews (e.g. github.com, gitlab.com). The simplified access to the contribution process attracted more developers than ever before. Barcomb et al. (Barcomb et al., 2018) defined a framework to understand the motives of peripheral contributors and the framework captured five critical aspects about the developers: motivation, social norms, community feeling, satisfaction and commitment.

Every contributor has his motivations which is a mixture of altruism and self-centered motives. Hyde et al. (Hyde et al., 2016) showed that both altruistic and self-centered motives were equally present among the contributors, but extrinsic motives like carrier opportunity or remuneration did not play a crucial role in retaining the contributors in the long run (Krishnamurthy et al., 2016). The newcomers remain with a project when they have intrinsic or altruistic motivations. The peripheral contributions were driven by the momentary relationship between the developer and the project, the general feeling of the developer towards the OSS projects did not play a role in these decisions. The volunteers kept contribute to the projects without remuneration because they were interested in and enjoyed, but their commitments were restricted by their daily job, family and available leisure time. The analyzed OSS projects lacked any practices to retain peripheral developers though predefined small tasks, bugs, translation or well-defined documentation tasks were suggested for newcomers to encourage the peripheral contributions. The predefined tasks help to solve a small, exciting problem immediately to prevent discouragement by missing technical skills. The easy, standardized contribution process (e.g. Github) simplified the submission and review processes while avoiding the necessary learning curve of a project-specific interface. The author used the sourceforge.net homepage to host the project homepage, source code repositories and downloadable software. The project source codes were open and released under a free software license. Both casual users and possible contributors could get the basic information on the project websites for their needs. Sourceforge has similar services for the easy contribution process like Github.

The effects of social norms were underestimated by the developers interviewed for the Barcomb framework. The episodic contributors did not make a link between their own social connections and OSS involvements, furthermore, they did not think their relatives or friends care about the volunteer activities. However, more episodic developers referred to somebody from their social connections who invited them to contribute to a project and this introduction method was the most frequent for non-code contributions. The past literature overlooked the power of social connections since they focused on the code contributions, but OSS can get better visibility and more contributions if the project members utilize their social networks. The version updates of AiBO+ project were announced on an online forum and Twitter. The engaged community members shared and retweeted these announcements in their social circle which could reach more potential users and contributors than the author alone.

According to the survey results in (Barcomb et al., 2018), the core OSS developers had no established practices for the peripheral developers and they associated the episodic contributions with low interest while most peripheral developers reported affinity for the OSS projects where they contributed. Although episodic developers tend to feel less attachment, the habitual developers cited more the community feeling like a reason why they contributed later again. The developers emphasized that the OSS communities are very tolerant and receptive to new members regardless of their geolocation, gender or sexual orientation. As the peripheral developers reported affinity to the OSS projects, the success of the AiBO+

project was driven by the enthusiastic robot dog owners in the online community. Since these customers owned their robots for years, they were in the technological acceptance phase and this passion helped an easier involvement in the collaborative AI development. The author also followed the general principle to accept the contributors regardless of their gender, religion, sexual orientation and location.

The satisfaction after contributions is triggered when the initial expectations are in line with the later feedback returned from the community. Similar to traditional volunteering, satisfaction was the best indicator to find those contributors who want to remain with OSS projects (Wu et al., 2007) thus it is important to keep the peripheral contributors satisfied with retentive practices. The survey results of Barcomb et al. confirmed that the peripheral developers are pleased to contribute to OSS projects when they feel appreciated, enjoy the contribution, help others and belong to a community. However, the core developers did not involve active strategies to foster the retention of episodic developers to become habitual. Although the source code contributions are usually thanked in some automated way in a "Contributors" file, the non-code submissions and bug reports are not honored anywhere. The author actively communicated with the contributors regarding the bug reports on the online forum and all submissions were publicly thanked in the software release announcements.

Bostrom reviewed the openness in the artificial intelligence development in a broader sense (Bostrom, 2017), openness of source code, science, data, safety techniques, capabilities and goals). He found that some forms of openness have positive effects like safety measures and goals, but others have negative implications (source code, science, capability) in this industry. He argued that these latter can cause a tougher competition to introduce a general AI because winning the AI race is incompatible with agreed safety measures, delays and performance limitations. The desired openness in AI development has trade-offs with open source codes, shard datasets and open algorithms.

The most famous AI developers apply a high level of openness. Big Western and Chinese companies with big AI departments present their latest achievements at conferences and upload their manuscripts to the ArXiv preprint deposit servers. Many AI publications are connected to released source codes or deep learning models to conduct reproducible research and make derivative works. Since these companies are in the interest of selling AI-related services, they wanted to accelerate AI research by releasing open source frameworks for AI researchers. On the other hand, other companies are less open and proprietary with their technologies because they are application-oriented and the competitive AI edge of these smaller companies is their bread and butter to attract venture capital. The author applied maximal openness during the collaborative development by free source codes, message boards, bug tracker and machine learning models.

After discussing the closest previous works in research and industry, the following paragraphs synthesize the questionnaire results, the community experiences and discussing these lessons in light of the current status of the social robotics industry.

Social Robotics Industry

A social robot must differentiate itself from other machines with unique skills, but it should not replace the functions of a smartphone or a computer, especially with more hassle. A balance should be maintained in the robot design between reducing the hardware costs and sacrificing basic skills otherwise consumers will not like the final product. The author found this issue within the AIBO community in Publication I. Although the Sony ERS-3xx had the lowest retail price among the Sony AIBO products, this price reduction implied software and hardware feature cuts. Despite the fact that this model had the most affordable price, the community members liked ERS-3xx the least because of the missed essential capabilities compared to more expensive AIBO models.

On the other hand, targeting a too complex robot design results in a long product life cycle and increased hardware costs that prevent any robot's mass adoption. With the current state-of-the-art technologies, a realistic product of the robotics companies should resemble a toy or an animal. As the uncanny valley showed in Figure 5, people accept a toy robot with a positive attitude since they do not expect high-level intelligence from the agent compared to a human-like object. Though a humanoid robot has a better acceptance, but the greater user expectations for the intelligence and the rising hardware costs make those robots inaccessible for wider consumer adoption. Sony AIBOs were successful because they resembled an animal, walked around and had interactions with humans according to the marketing promises. The robot owners expected less intelligence from Sony AIBOs than a humanoid robot and the robot appearance was in line with its capabilities to avoid the uncanny valley.

Pleo owners were disappointed in (Fernaeus et al., 2010) when they realized that their robots have legs, but they cannot walk autonomously. Cozmo and Vector robots by Anki were good examples from the industry to keep the hardware costs at a relatively low level while satisfying the consumer expectations. Over 200000 units were sold from these robots (Techcrunch, 2018) with the market prices between $150-250 and they could gather a quite large developer community of over 15000 developers (Palatucci, 2018). Although the first three generations of Sony AIBO robots were sold over 175000 units (“AIBO,” 2018) despite their >$1000 market prices, but those robots were more complex compared to the Anki products. The lack of a modern software development environment limited the third-party programmers in the AIBO community. These assumptions were confirmed for the AiBO+ project, since, despite AIBO robots were good products, peripheral contributions were only received from the community because of the technical difficulties. The author got know multiple community members who were supposed to contribute code, but they gave up after realizing how difficult the development of AIBO robots. Looking at the community involvement in Nao development (see Section 2.3), the same phenomenon can be identified. Aldebaran Robotics had the vision to work with their customers together, they dedicated internal people to deal with the 3rd party development and they open-sourced some parts of their software stack. They received a considerable amount of contributions over the years, but they were not satisfied with the achievements. As their robot was expensive, it resembled a short human and working with bipedal robots is challenging, they attracted technical people. Their forum and developer community composed of mainly professional developers, students and researchers. The future projects need to put attention for a low technical barrier in order to get non-technical habitual or core contributions.

When robotics companies enter the market, unpredictable obstacles can appear in product development. This challenge is not limited to the special market entry with crowdfunding, but robot hardware and AI development have an inherent, high complexity. A silver bullet does not exist to solve this situation. Transparent product development and clear communication about the software releases and any delays to the customers show about the company that it cares about the target audience.

The demographics of the participants, who answered the questionnaire (see Section 3.3.1), revealed that the social robots can be targeted from the teenagers to the pensioners. Furthermore, roughly half of the owners were older than 40 years and 30 % of them were female. This survey was run long after the AIBO robots were discontinued from the market, but a study (Fujita, 2004) found similar proportions for the AIBO customers in 2002. Therefore, the age and gender distributions of the owners have been remained stable from the beginning until the sales were stopped and these robots were traded on the secondhand market. Another finding was the large fraction of non-technical owners. If these people are considered for community contributions, the project must plan appropriate tasks which have a quick learning curve. Non-technical contributions often require little technical knowledge, but more creativity, imagination and patience. The author experienced in AiBO+ that the community had a self-organized innovation to solve upcoming problems. It can be expected that all online communities have this internal property and there is a high chance to create successful strategies from the realization of community ideas.

When the capabilities of a robot are presented in a promotional video, often eye-catching AI features are demonstrated, but the development teams struggle to turn these promises into reality. While the AI development is challenging in general, the questionnaire results showed that the robots must adapt their personalities according to the human gender, age and culture to maximize their success in these diverse target groups. A good example was when older participants were compared to younger, they treated the robots as a toy, they were less positive about the social skills of the robot, but they were keen to create content contributions. In general, the interaction and conversational skills were found the most important by the robot owners, this perceived sociability was a key point to accept the robot.

We live in a world where technology is everywhere around us and science fiction movies portray the agents with artificial intelligence which can let us reaching various services. The people expect these perspectives of the AI in the present robots. Since many internet services and conversational agents are easily accessible in the smartphones, the integration of these features was a demand from the heavy users of Sony AIBO. Thus, the long-term ownership did not degrade the robot acceptance after several years, but the owners wished to integrate the latest available technologies. The AIBO owners also desired the learning and memory skills although these are hard AI problems, they can be prioritized less in the development process than the interaction skills. Another depiction of the robots in storytelling is that they are self-conscious and can move around independently. When a new mobile robot is promoted with wheels or legs, the customers expect autonomy and self-charging. Their excitement turns into disappointment if they find out, the robot does not move according to their assumptions.

As the author described earlier, the complexity of artificial intelligence encumbers the development of new social robots. A solution can be the collaboration with the community around the product to enhance the AI skills to meet the high expectations over time. The questionnaire answers showed that the robot owners got bored with the repetitive behaviors and expected new content. However, they did not abandon their robot because the original AI software built an emotional bond between the robots and the owners. To improve this situation, this opportunity can be turned into a business strategy. The robot consumers can be charged a fee for AI updates, similar to the mobile applications. A robotics company can reach sustainable revenue with this approach.

The general principles for robotics startups can be summarized in some points:

- Define the robot's target audience and justify the utility value over conventional devices (e.g. computers, handsets, smart speakers). Run a survey among the target audience to gather their expectations for the product. This knowledge is inevitable to have an accomplishable development roadmap.

- Robots have complex hardware and software while the customer expectations are high due to the depictions in the science fiction stories. These products cannot be treated as a usual consumer device in the product planning and execution.
- AI development has hidden obstacles regardless of the best planning. Clear communication to the customers about the delays enhances the trust and the future community around the robot.
- If the robot does not have a direct utility (e.g. housekeeping tasks), plan a robot which keeps the expectations low. For example, aesthetics with toy or animal appearance.
- Create a communication channel (e.g. forum) where questions regarding the robot can be answered and the customers can talk to each other, even before the market entry. Dedicate people to talk with the customers on the forum which is a separate role compared to the customer services.
- Define minimal, feasible skills for the AI engine which can be developed in a reasonable timeframe until the product release. The artificial intelligence features must be prioritized and implemented in priority order.
- Think about the subtle differences in the robot behaviors inside the target audience. There can be gender, cultural and other modulations in the people's expectations.
- Execute a plan to involve the community in continuous development after the initial product is released.
- Expect that a large proportion of the community does not have technical skills, but they still would like to collaborate in AI development. These people will not become core contributors for the project rather habitual or episodic contributors.
- Consider distributing the community contributions in free content updates, but create a subscription plan for the owners to get AI updates from the core development team. This approach has the potential to provide continuous revenue for the company after receiving the initial sale price while keeping community contributions royalty-free.

3.8 SUMMARY

This chapter looked through the collaborative development project AiBO+ which was executed by the author. The whole timeline and technical aspects were discussed. The analysis of a questionnaire was presented with two articles (Publication I and II) to identify the properties of an online community that consisted of heavy users of Sony AIBO robots. Their attitudes towards their robots were identified, their wishes for new content updates and their intentions for providing contributions. Three other articles (Publication III, IV and V) described some technical details of the initial content offering to get the community's initial attention and engage the member for contribution over time. At the end of the chapter, the received community contributions were reviewed and best practices were listed for future robotics projects. The final chapters will discuss the resolution of the research questions and draw the final conclusions.

4 Conclusions

The following research questions were explored in the dissertation:

RQ1: How long-term owners of a social robot can be involved in collaborative artificial intelligence development?

Although the OSS practices can be applied to some extent, the artificial intelligence engines are complex, special software thus the community engagement might require different strategies. The author approached the problem by joining an online forum with long-term Sony AIBO robot owners and building credibility among the community members. The author participated in the discussions and helped to solve the problems of other members. Some ERS-7 robots were bought for the project, an initial content offering was developed and released to the community for free. When any bugs were discovered in the AI engine of the project, the author fixed them quickly. This proactiveness gave the community members trust that the author will run this project for a long time. Since the robot owners had a positive experience with the initial releases, several AIBO community members were engaged for episodic contributions. These contributions had a high value as the members were not remunerated for their work.

This achievement underscored the importance of the appropriate social robot design. When the owners have an emotional connection with their robots, they can be engaged in collaborative AI development more efficiently. This connection can be built with the right aesthetics and enjoyable initial AI skills to avoid the uncanny valley. Another aspect is the communication language during the collaborative development. The forum language was English and the responders to the questionnaire stated they have good English knowledge. The author did not experience any misunderstandings during the collaborative development, but the English language can restrict the potential contributions. The author must give a positive, gentle pressure few times to get the final contributions with engaged members and this approach always worked. If somebody changed their mind about contributing the project, the author let the opportunity go without any aggressive strategies. In general, code, content and scientific contributions were received from episodic developers, the AiBO+ project was a success.

RQ2: How can future social robotics projects run collaborative artificial intelligence development successfully?

The review of the OSS practices and the industry given valuable insights about the current status in the social robotics. After the dissertation project was finished, several findings were discovered to improve the success rate of future social robotics projects. First of all, robots cannot be treated like other products on the market because they are special caused by their costs and complexity. This misconception led to the fall of the Jibo company. Thus the most important discovery was the consequences of the robot design. The aesthetics of a robot determinates the expectations of the future owners, the hardware expenses and the position on the uncanny valley (Figure 5). The more similar a robot to a human, the harder to develop a matching AI because of the increased expectations from the consumers. However, robotics hardware with a matching AI can make an emotional bond with their owners and be part of their lives. It was found that more robotics companies had certain strategies to communicate with the communities of their robots and initiate collaborative development. However, these efforts concentrated mainly on professional coders to get code contributions, but the majority of the robot customers are non-technical people. To get non-technical contributions, the companies must aim a short learning curve for their development environment and define such contributions that fit into their AI and they are easy to create. The subscription-based monthly fees also seem to be an unavoidable business strategy to get regular funding for the continuous AI development for the products. More detailed recommendations are in Section 3.7.

In summary, this dissertation presented a research topic of collaborative artificial intelligence development. Two publications focused on understanding an online community of social robot owners to execute the collaborative development project and three articles detailed some technical aspects of the initial content to attract members for contribution.

The work had multiple scientific contributions. The first is the analysis of an online community of social robot owners by a questionnaire who possessed and used their robots long after they were discontinued from the market. The heavy users of social robots with years-long ownership were never studied before in the robotics field. All long-term experiments with robots lasted no more than a few months in the past and the experiments were usually carried out in labs. Unlike the robot owners in this research operated their robots inside their homes. The main focus of the questionnaire was the robot acceptance after years-long robot ownership and the consequences to the participation in a collaborative AI development. Therefore, the aforementioned online community was analyzed from multiple points of view. The author examined some variables (e.g. age, gender) of the community population which were subjects of earlier works in the literature. This way the old findings could be reexamined with heavy users. The questionnaire evaluation gave a few recommendations for social robot design regarding the earlier by other researchers in Publication I and II. These discoveries were complemented with the results from the review of the social robotics industry and they were turned to generic recommendations for social robotics companies in Section 3.7. This research had another unique circumstance because the author run this case study of collaborative AI development as an insider in the online community and the progress of the AiBO+ project evolved on its own over the years.

Furthermore, the collaborative artificial intelligence development was never examined in the past, the research literature targeted open-source projects, community-based product innovations and designs. The author reviewed the current state of the robotics industry and the existing community practices in the companies. The findings were compared to the results of the studies dealing with OSS projects.

Technical results of the scientific outcomes included three original articles for sound event, body state and surface recognitions based on machine learning methods. Moreover, pursuing the functional correctness of the AI software, the deterministic test cases were developed for behavior-based robotics to ensure the reproducibility and quick testing for robot software.

4.1 FUTURE WORK

Since this doctoral dissertation was an exploratory work, there are several options to extend this knowledge. The same collaborative development principles can be tried in an online community of another social robot or a similar questionnaire can examine robot acceptance among the heavy users of other robots like Cozmo or Vector. The collaborative development can be studied in other robotics companies to get a better understanding of the industry with newer insights and finetune the recommendations of this dissertation. Eventually, the non-technical contributions to AI development can be extended further beyond the findings of this dissertation. The close cooperation with the target audience and an understanding of its motives via questionnaire can set a role model for future projects in the fast-changing scenario of social robots for healthcare, service and educational applications.

The questionnaire results correspond to the current state of society and the technological environment. The target group's evaluation should be repeated in the future since the user expectations change over time.

5 References

Ackerman, E. (2017, January 3). *Mayfield Robotics Announces Kuri, a $700 Home Robot*. IEEE Spectrum. https://spectrum.ieee.org/mayfield-robotics-announces-kuri-a-700-mobile-home-robot

AiBO+ Client (3.1.2). (2020). [Computer software]. https://apps.apple.com/in/app/aibo-client-for-sony-ers-7-robots/id1105335533 and https://play.google.com/store/apps/details?id=fi.tampere.aiboplus.mindclientapp

AIBO. (2018). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=AIBO&oldid=874573585

Ayesh, A., Joseph, C., Perril, S., & Thomas, S. (2014). Aesthetics of a robot: Case study on aibo dog robots for buddy-ing devices. *Journal of Intelligent Computing*, *5*(1), 1–15.

Bagozzi, R. P., & Dholakia, U. M. (2006). Open Source Software User Communities: A Study of Participation in Linux User Groups. *Management Science*, *52*, 1099–1115.

Bajcsy, A., Losey, D. P., O'Malley, M. K., & Dragan, A. D. (2017). Learning robot objectives from physical human interaction. *Proceedings of Machine Learning Research*, *78*, 217–226.

Barcomb, A., Kaufmann, A., Riehle, D., Stol, K.-J., & Fitzgerald, B. (2018). Uncovering the periphery: A qualitative survey of episodic volunteering in free/libre and open source software communities. *IEEE Transactions on Software Engineering*.

Barnes, J., FakhrHosseini, M., Jeon, M., Park, C.-H., & Howard, A. (2017). The influence of robot design on acceptance of social robots. *2017 14th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI)*, 51–55.

Bartneck, C., Suzuki, T., Kanda, T., & Nomura, T. (2007). The influence of people's culture and prior experiences with Aibo on their attitude towards robots. *AI & SOCIETY*, *21*(1), 217–230. https://doi.org/10.1007/s00146-006-0052-7

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., & Tanaka, F. (2018). Social robots for education: A review. *Science Robotics*, *3*(21).

Björling, E. A., Rose, E., Davidson, A., Ren, R., & Wong, D. (2020). Can we keep him forever? Teens' engagement and desire for emotional connection with a social robot. *International Journal of Social Robotics*, *12*(1), 65–77.

Bostrom, N. (2017). Strategic implications of openness in AI development. *Global Policy*, *8*(2), 135–148.

Bradley, M. M., Codispoti, M., Sabatinelli, D., & Lang, P. J. (2001). Emotion and motivation II: Sex differences in picture processing. *Emotion (Washington, D.C.)*, *1*(3), 300–319.

Breazeal, C. (2017). Social robots: From research to commercialization. *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, 1–1.

Brody, L. R. (1997). Gender and Emotion: Beyond Stereotypes. *Journal of Social Issues*, *53*(2), 369–393. https://doi.org/10.1111/j.1540-4560.1997.tb02448.x

Burget, F., Fiederer, L. D. J., Kuhner, D., Völker, M., Aldinger, J., Schirrmeister, R. T., Do, C., Boedecker, J., Nebel, B., & Ball, T. (2017). Acting thoughts: Towards a mobile robotic service assistant for users with limited communication skills. *2017 European Conference on Mobile Robots (ECMR)*, 1–6.

Bussgang, J. J., & Snively, C. (2015). Jibo: A Social Robot for the Home. *Harvard Business School Case 816-003*.

Cameron, D., Fernando, S., Collins, E., Millings, A., Moore, R., Sharkey, A., Evers, V., & Prescott, T. (2015). Presence of life-like robot expressions influences children's enjoyment of human-robot interactions in the field. *Proceedings of the AISB Convention 2015*.

Carillo, K., Huff, S., & Chawner, B. (2017). What makes a good contributor? Understanding contributor behavior within large Free/Open Source Software projects A socialization perspective. *J. Strateg. Inf. Syst.*, *26*(4), 322–359.

Castro-González, Á., Castillo, J. C., Alonso-Martín, F., Olortegui-Ortega, O. V., González-Pacheco, V., Malfaz, M., & Salichs, M. A. (2016). The effects of an impolite vs. A polite robot playing rock-paper-scissors. *International Conference on Social Robotics*, 306–316.

Chan, L., Zhang, B. J., & Fitter, N. T. (2021). Designing and Validating Expressive Cozmo Behaviors for Accurately Conveying Emotions. *2021 30th IEEE International Conference on Robot Human Interactive Communication (RO-MAN)*, 1037–1044. https://doi.org/10.1109/RO-MAN50785.2021.9515425

Churamani, N., Kerzel, M., Strahl, E., Barros, P., & Wermter, S. (2017). Teaching emotion expressions to a human companion robot using deep neural architectures. *2017 International Joint Conference on Neural Networks (IJCNN)*, 627–634.

Cormier, D., Newman, G., Nakane, M., Young, J. E., & Durocher, S. (2013). Would you do as a robot commands? An obedience study for human-robot interaction. *International Conference on Human-Agent Interaction*.

Coviello, N. E., & Joseph, R. M. (2012). Creating major innovations with customers: Insights from small and young technology firms. *Journal of Marketing*, *76*(6), 87–104.

Crowston, K., Wei, K., Howison, J., & Wiggins, A. (2012). Free/Libre open-source software development: What we know and what we do not know. *ACM Computing Surveys*, *44*(2), 1–35. https://doi.org/10.1145/2089125.2089127

da Mota Pedrosa, A. (2012). Customer integration during innovation development: An exploratory study in the logistics service industry. *Creativity and Innovation Management*, *21*(3), 263–276.

de Graaf, M. M., Allouch, S. B., & van Dijk, J. A. (2016). Long-term acceptance of social robots in domestic environments: Insights from a user's perspective. *2016 AAAI Spring Symposium Series*.

de Graaf, M. M., Ben Allouch, S., & van Dijk, J. A. (2019). Why would I use this in my home? A model of domestic social robot acceptance. *Human–Computer Interaction*, *34*(2), 115–173.

Deng, E., Mutlu, B., & Mataric, M. (2019). Embodiment in socially interactive robots. *ArXiv Preprint ArXiv:1912.00312*.

Dinh-Trong, T. T., & Bieman, J. M. (2005). The FreeBSD Project: A Replication Case Study of Open Source Development. *IEEE Trans. Softw. Eng.*, *31*(6), 481–494.

Doering, M., Glas, D. F., & Ishiguro, H. (2019). Modeling interaction structure for robot imitation learning of human social behavior. *IEEE Transactions on Human-Machine Systems*, *49*(3), 219–231.

Edwards, A., Edwards, C., Westerman, D., & Spence, P. R. (2019). Initial expectations, interactions, and beyond with social robots. *Computers in Human Behavior*, *90*, 308–314.

Émond, C., Lewis, L., Chalghoumi, H., & Mignerat, M. (2020). A Comparison of NAO and Jibo in Child-Robot Interaction. *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 192–194. https://doi.org/10.1145/3371382.3378234

Ezer, N. (2008). *Is a robot an appliance, teammate, or friend? Age-related differences in expectations of and attitudes towards personal home-based robots*. Georgia Institute of Technology.

Fang, E. (2008). Customer participation and the trade-off between new product innovativeness and speed to market. *Journal of Marketing*, *72*(4), 90–104.

Ferland, F., Létourneau, D., Aumont, A., Frémy, J., Legault, M.-A., Lauria, M., & Michaud, F. (2013). Natural interaction design of a humanoid robot. *Journal of Human-Robot Interaction*, *1*(2), 118–134.

Fernaeus, Y., H\a akansson, M., Jacobsson, M., & Ljungblad, S. (2010). How Do You Play with a Robotic Toy Animal?: A Long-term Study of Pleo. *Proceedings of the 9th International Conference on Interaction Design and Children*, 39–48. https://doi.org/10.1145/1810543.1810549

Fink, J., Mubin, O., Kaplan, F., & Dillenbourg, P. (2012). Anthropomorphic language in online forums about Roomba, AIBO and the iPad. *2012 IEEE Workshop on Advanced Robotics and Its Social Impacts (ARSO)*, 54–59. https://doi.org/10.1109/ARSO.2012.6213399

Fujita, M. (2000). Entertainment robot: AIBO. *The Journal of the Institute of Image Information and Television Engineers*, *54*(5), 657–661.

Fujita, M. (2004). On activating human communications with pet-type robot AIBO. *Proceedings of the IEEE*, *92*(11), 1804–1813. https://doi.org/10.1109/JPROC.2004.835364

Fujita, M., Kitano, H., & Doi, T. (2000). Robot entertainment. *Robots for Kids: Exploring New Technologies for Learning*, 37–72.

Garcia, A., Sant'Anna, C., Figueiredo, E., Kulesza, U., Lucena, C., & Staa, A. (2005). Modularizing design patterns with aspects: A quantitative study. *Aspect-Oriented Software Development: Proceedings of the 4 Th International Conference on Aspect-Oriented Software Development*, *14*, 3.

Garza, C. G. M. (2018). *Failure Is an Option: How the Severity of Robot Errors Affects Human-Robot Interaction*. Carnegie Mellon University Pittsburgh, PA.

Geva, N., Uzefovsky, F., & Levy-Tzedek, S. (2020). Touching the social robot PARO reduces pain perception and salivary oxytocin levels. *Scientific Reports*, *10*(1), 1–15.

Ghaleb, T., Aljasser, K., & Alturki, M. (2015). *Implementing the Observer Design Pattern as an Expressive Language Construct*. https://doi.org/10.13140/RG.2.1.4879.8164

Gousios, G., Pinzger, M., & Deursen, A. van. (2014). An exploratory study of the pull-based software development model. *Proceedings of the 36th International Conference on Software Engineering*, 345–355.

Graaf, D. (2015). *Living with robots: Investigating the user acceptance of social robots in domestic environments*.

Graaf, M. M. A. de, Allouch, S. B., & Dijk, J. A. G. M. van. (2014, April 3). *Long-term evaluation of a social robot in real homes*. 3rd International Symposium on New Frontiers in Human-Robot Interaction 2014: A two-day Symposium at AISB 2014 Convention. https://research.utwente.nl/en/publications/long-term-evaluation-of-a-social-robot-in-real-homes

Greer, C. R., & Lei, D. (2012). Collaborative innovation with customers: A review of the literature and suggestions for future research. *International Journal of Management Reviews*, *14*(1), 63–84.

Haring, K. S., Matsumoto, Y., & Watanabe, K. (2013). How do people perceive and trust a lifelike robot. *Proceedings of the World Congress on Engineering and Computer Science*, *1*.

Haring, K. S., Mougenot, C., Ono, F., & Watanabe, K. (2014). Cultural differences in perception and attitude towards robots. *International Journal of Affective Engineering*, *13*(3), 149–157.

Haring, K. S., Silvera-Tawil, D., Takahashi, T., Velonaki, M., & Watanabe, K. (2015). Perception of a humanoid robot: A cross-cultural comparison. *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 821–826.

Heerink, M., Krose, B., Evers, V., & Wielinga, B. (2006). The influence of a robot's social abilities on acceptance by elderly users. *ROMAN 2006-The 15th IEEE International Symposium on Robot and Human Interactive Communication*, 521–526.

Heerink, M., Kröse, B., Evers, V., & Wielinga, B. (2008). *The influence of social presence on acceptance of a companion robot by older people*.

Heerink, M., Kröse, B., Evers, V., & Wielinga, B. (2010). Relating conversational expressiveness to social presence and acceptance of an assistive social robot. *Virtual Reality*, *14*(1), 77–84.

Hemetsberger, A., & Godula, G. (2007). Integrating expert customers in new product development in industrial business-virtual routes to success. *Innovative Marketing*, *3*(3), 28–39.

Hempel, E., Fischer, H., Gumb, L., Höhn, T., Krause, H., Voges, U., Breitwieser, H., Gutmann, B., Durke, J., & Bock, M. (2003). An MRI-compatible surgical robot for precise radiological interventions. *Computer Aided Surgery*, *8*(4), 180–191.

Hendricks, K. B., & Singhal, V. R. (1997). Delays in new product introductions and the market value of the firm: The consequences of being late to the market. *Management Science*, *43*(4), 422–436.

Hiroi, Y., & Ito, A. (2013). ASAHI: OK for failure A robot for supporting daily life, equipped with a robot avatar. *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 141–142.

Hockstein, N. G., Gourin, C. G., Faust, R. A., & Terris, D. J. (2007). A history of robots: From science fiction to surgical robotics. *Journal of Robotic Surgery*, *1*(2), 113–118.

Horstmann, A. C., & Krämer, N. C. (2020). Expectations vs. Actual behavior of a social robot: An experimental investigation of the effects of a social robot's interaction skill level and its expected future role on people's evaluations. *Plos One*, *15*(8), e0238133.

Huh, M., Agrawal, P., & Efros, A. A. (2016). What makes ImageNet good for transfer learning? *ArXiv:1608.08614 [Cs]*. http://arxiv.org/abs/1608.08614

Hung, L., Liu, C., Woldum, E., Au-Yeung, A., Berndt, A., Wallsworth, C., Horne, N., Gregorio, M., Mann, J., & Chaudhury, H. (2019). The benefits of and barriers to using a social robot PARO in care settings: A scoping review. *BMC Geriatrics*, *19*(1), 232.

Hyde, M. K., Dunn, J., Bax, C., & Chambers, S. K. (2016). Episodic Volunteering and Retention:An Integrated Theoretical Approach. *Nonprofit and Voluntary Sector Quarterly*, *45*(1), 45–63.

Innes, J., & Morrison, B. (2017). *Projecting the future impact of advanced technologies: Will a robot take my job?* https://researchoutput.csu.edu.au/en/publications/projecting-the-future-impact-of-advanced-technologies-will-a-robo

Jensen, L. S., & Özkil, A. G. (2018). Identifying challenges in crowdfunded product development: A review of Kickstarter projects. *Design Science*, *4*.

Jones, J. L. (2006). Robots at the tipping point: The road to iRobot Roomba. *IEEE Robotics Automation Magazine*, *13*(1), 76–78. https://doi.org/10.1109/MRA.2006.1598056

Jung, M., & Hinds, P. (2018). *Robots in the wild: A time for more robust theories of human-robot interaction*. ACM New York, NY, USA.

Kahn, P. H., Friedman, B., & Hagman, J. (2002). "I care about him as a pal": Conceptions of robotic pets in online AIBO discussion forums. *CHI '02 Extended Abstracts on Human Factors in Computing Systems*, 632–633. https://doi.org/10.1145/506443.506519

Kaplan, F. (2004). Who is Afraid of the Humanoid? Investigating Cultural Differences in the Acceptance of Robots. *I. J. Humanoid Robotics*, *1*, 465–480. https://doi.org/10.1142/S0219843604000289

Kennedy, J., Lemaignan, S., Montassier, C., Lavalade, P., Irfan, B., Papadopoulos, F., Senft, E., & Belpaeme, T. (2017). Child speech recognition in human-robot interaction: Evaluations and recommendations. *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, 82–90.

Kertész, C. (2013). Improvements in the native development environment for Sony AIBO. *International Journal of Interactive Multimedia and Artificial Intelligence*, *2*(3 (Special Issue on Improvements in Information Systems and Technologies)), 50–54.

Kertesz, C., & Turunen, M. (2017). *Reproducible Testing for Behavior-Based Robotics*. 1–3.

Kindler, A., Golosovsky, M., & Solomon, S. (2019). Early prediction of the outcome of Kickstarter campaigns: Is the success due to virality? *Palgrave Communications*, *5*(1), 1–6. https://doi.org/10.1057/s41599-019-0261-6

Kostavelis, I., Giakoumis, D., Malassiotis, S., & Tzovaras, D. (2016). Human aware robot navigation in semantically annotated domestic environments. *International Conference on Universal Access in Human-Computer Interaction*, 414–423.

Krishnamurthy, R., Jacob, V., Radhakrishnan, S., & Dogan, K. (2016). Peripheral Developer Participation in Open Source Projects: An Empirical Analysis. *ACM Trans. Manage. Inf. Syst.*, *6*(4), 1–31.

Lasschuijt, M. (2019). *Communication Style In Online Crowdfunding*. https://www.semanticscholar.org/paper/Communication-Style-In-Online-Crowdfunding-Lasschuijt/5951e00d0860c082f43dc2ed4c4c8a66c0282859

Leite, I., Martinho, C., & Paiva, A. (2013). Social Robots for Long-Term Interaction: A Survey. *International Journal of Social Robotics*, *5*(2), 291–308.

Leite, I., Pereira, A., Mascarenhas, S., Martinho, C., Prada, R., & Paiva, A. (2013). The influence of empathy in human–robot relations. *International Journal of Human-Computer Studies*, *71*(3), 250–260.

Lemaignan, S., Warnier, M., Sisbot, E. A., Clodic, A., & Alami, R. (2017). Artificial cognition for social human–robot interaction: An implementation. *Artificial Intelligence*, *247*, 45–69.

Liu, P., Glas, D. F., Kanda, T., & Ishiguro, H. (2016). Data-driven HRI: Learning social behaviors by example from human–human interaction. *IEEE Transactions on Robotics*, *32*(4), 988–1008.

Liu, R., & Zhang, X. (2019). A review of methodologies for natural-language-facilitated human–robot cooperation. *International Journal of Advanced Robotic Systems*, *16*(3), 1729881419851402.

March, S. T., & Smith, G. F. (1995). Design and natural science research on information technology. *Decision Support Systems*, *15*(4), 251–266. https://doi.org/10.1016/0167-9236(94)00041-2

McAlexander, J. H., Schouten, J. W., & Koenig, H. F. (2002). Building brand community. *Journal of Marketing*, *66*(1), 38–54.

Miklósi, Á., Korondi, P., Matellán, V., & Gácsi, M. (2017). Ethorobotics: A new approach to human-robot relationship. *Frontiers in Psychology*, *8*, 958.

Mirnig, N., Stadler, S., Stollnberger, G., Giuliani, M., & Tscheligi, M. (2016). Robot humor: How self-irony and Schadenfreude influence people's rating of robot likability. *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 166–171.

Mirnig, N., Stollnberger, G., Miksch, M., Stadler, S., Giuliani, M., & Tscheligi, M. (2017). To err is robot: How humans assess and act toward an erroneous social robot. *Frontiers in Robotics and AI*, *4*, 21.

Mori, M. (1970). The uncanny valley. *Energy*, *7*(4), 33–35.

Mubin, O., Wadibhasme, K., Jordan, P., & Obaid, M. (2019). Reflecting on the Presence of Science Fiction Robots in Computing Literature. *ACM Transactions on Human-Robot Interaction*, *8*(1), 5:1-5:25. https://doi.org/10.1145/3303706

Nagle, F. (2017). *Learning by Contributing: Gaining Competitive Advantage Through Contribution to Crowdsourced Public Goods* (SSRN Scholarly Paper ID 3091831). Social Science Research Network. https://papers.ssrn.com/abstract=3091831

Nauta, J., Mahieu, C., Michiels, C., Ongenae, F., De Backere, F., De Turck, F., Khaluf, Y., & Simoens, P. (2019). Pro-active positioning of a social robot intervening upon behavioral disturbances of persons with dementia in a smart nursing home. *Cognitive Systems Research*, *57*, 160–174.

Nomura, T., Sugimoto, K., Syrdal, D. S., & Dautenhahn, K. (2012). Social acceptance of humanoid robots in Japan: A survey for development of the frankenstein syndorome questionnaire. *2012 12th IEEE-RAS International Conference on Humanoid Robots (Humanoids 2012)*, 242–247. https://doi.org/10.1109/HUMANOIDS.2012.6651527

Nomura, T., Suzuki, T., & Kanda, T. (2006). *Altered Attitudes of People toward Robots: Investigation through the Negative Attitudes toward Robots Scale **.

Nomura, T., Tasaki, T., Kanda, T., Shiomi, M., Ishiguro, H., & Hagita, N. (2005). Questionnaire–Based Research on Opinions of Visitors for Communication Robots at an Exhibition in Japan. In M. F. Costabile & F. Paternò (Eds.), *Human-Computer Interaction – INTERACT 2005* (pp. 685–698). Springer Berlin Heidelberg.

Obaid, M., Sandoval, E. B., Złotowski, J., Moltchanova, E., Basedow, C. A., & Bartneck, C. (2016). Stop! That is close enough. How body postures influence human-robot proximity. *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 354–361.

Palatucci, M. (2018, September 17). *Cozmo*. CITRIS People and Robots Initiative, UC Berkeley. https://www.youtube.com/watch?v=hfP0E7hZbMQ

Pinter, M., Lai, F., Sanchez, D. S., Ballantyne, J., Roe, D. B., Wang, Y., Jordan, C. S., Taka, O., & Wong, C. W. (2015). *Social behavior rules for a medical telepresence robot*. Google Patents.

Pinto, G., Steinmacher, I., & Gerosa, M. A. (2016). More common than you think: An in-depth study of casual contributors. *2016 IEEE 23rd International Conference on Software Analysis, Evolution, and Reengineering (SANER)*, *1*, 112–123.

Poulsen, A., Burmeister, O. K., & Kreps, D. (2018). The ethics of inherent trust in care robots for the elderly. *IFIP International Conference on Human Choice and Computers*, 314–328.

Prassler, E., Ritter, A., Schaeffer, C., & Fiorini, P. (2000). A short history of cleaning robots. *Autonomous Robots*, *9*(3), 211–226.

Rantanen, P., Parkkari, T., Leikola, S., Airaksinen, M., & Lyles, A. (2017). An in-home advanced robotic system to manage elderly home-care patients' medications: A pilot safety and usability study. *Clinical Therapeutics*, *39*(5), 1054–1061.

Riek, L. D. (2016). Robotics technology in mental health care. In *Artificial intelligence in behavioral and mental health care* (pp. 185–203). Elsevier.

Rossi, S., Ferland, F., & Tapus, A. (2017). User profiling and behavioral adaptation for HRI: a survey. *Pattern Recognition Letters*, *99*, 3–12.

Russell, S., & Norvig, P. (2009). *Artificial Intelligence: A Modern Approach* (3rd edition). Pearson.

Šabanović, S. (2010). Robots in Society, Society in Robots. *International Journal of Social Robotics*, *2*(4), 439–450. https://doi.org/10.1007/s12369-010-0066-7

Šabanović, S. (2014). Inventing Japan's 'robotics culture': The repeated assembly of science, technology, and culture in social robotics. *Social Studies of Science*, *44*(3), 342–367. https://doi.org/10.1177/0306312713509704

Šabanović, S., Bennett, C. C., Chang, W.-L., & Huber, L. (2013). PARO robot affects diverse interaction modalities in group sensory therapy for older adults with dementia. *2013 IEEE 13th International Conference on Rehabilitation Robotics (ICORR)*, 1–6.

Saffari, E., Hosseini, S. R., Taheri, A., & Meghdari, A. (2021). "Does cinema form the future of robotics?": A survey on fictional robots in sci-fi movies. *SN Applied Sciences*, *3*(6), 655. https://doi.org/10.1007/s42452-021-04653-x

Salichs, M. A., Barber, R., Khamis, A. M., Malfaz, M., Gorostiza, J. F., Pacheco, R., Rivas, R., Corrales, A., Delgado, E., & García, D. (2006). Maggie: A robotic platform for human-robot social interaction. *2006 IEEE Conference on Robotics, Automation and Mechatronics*, 1–7.

Sandmeier, P. (2009). Customer integration strategies for innovation projects: Anticipation and brokering. *International Journal of Technology Management*, *48*(1), 1–23.

Scopelliti, M., Giuliani, M. V., & Fornara, F. (2005). Robots in a domestic setting: A psychological approach. *Universal Access in the Information Society*, *4*(2), 146–155.

Seibt, J., Vestergaard, C., & Damholdt, M. F. (2021). The Complexity of Human Social Interactions Calls for Mixed Methods in HRI: Comment on "A Primer for Conducting Experiments in Human-robot Interaction," by G. Hoffman and X. Zhao. *ACM Transactions on Human-Robot Interaction*, *10*(1), 1–4. https://doi.org/10.1145/3439715

Seidlitz, L., & Diener, E. (1998). Sex differences in the recall of affective experiences. *Journal of Personality and Social Psychology*, *74*(1), 262–271. https://doi.org/10.1037//0022-3514.74.1.262

Stencel, K., & Węgrzynowicz, P. (2008). Implementation Variants of the Singleton Design Pattern. In R. Meersman, Z. Tari, & P. Herrero (Eds.), *On the Move to Meaningful Internet Systems: OTM 2008 Workshops* (pp. 396–406). Springer. https://doi.org/10.1007/978-3-540-88875-8_61

Sünderhauf, N., Brock, O., Scheirer, W., Hadsell, R., Fox, D., Leitner, J., Upcroft, B., Abbeel, P., Burgard, W., Milford, M., & Corke, P. (2018). The limits and potentials of deep learning for robotics. *The International Journal of Robotics Research*, *37*(4–5), 405–420. https://doi.org/10.1177/0278364918770733

Takahashi, M., Suzuki, T., Shitamoto, H., Moriguchi, T., & Yoshida, K. (2010). Developing a mobile robot for transport applications in the hospital domain. *Robotics and Autonomous Systems*, *58*(7), 889–899.

Talebpour, Z., & Martinoli, A. (2018). Multi-robot coordination in dynamic environments shared with humans. *2018 IEEE International Conference on Robotics and Automation (ICRA)*, 1–8.

Techcrunch. (2018, July). Anki has sold 1.5 million robots. *TechCrunch*. http://social.techcrunch.com/2018/08/08/anki-has-sold-1-5-million-cozmo-robots/

Thimmesch-Gill, Z., Harder, K. A., & Koutstaal, W. (2017). Perceiving emotions in robot body language: Acute stress heightens sensitivity to negativity while attenuating sensitivity to arousal. *Computers in Human Behavior*, *76*, 59–67.

Trivedi, K., Trivedi, P., & Goswami, V. (2018). Sustainable marketing strategies: Creating business value by meeting consumer expectation. *Undefined*. https://www.semanticscholar.org/paper/Sustainable-marketing-strategies%3A-Creating-business-Trivedi-Trivedi/efce487be9b2e20bb5518652f23e47419ca1972a

Trovato, G., Paredes, R., Balvin, J., Cuellar, F., Thomsen, N. B., Bech, S., & Tan, Z.-H. (2018). The sound or silence: Investigating the influence of robot noise on proxemics. *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 713–718.

von der Pütten, A., & Krämer, N. (2012). A survey on robot appearances. *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 267–268.

Vu, C., Cross, M., Bickmore, T., Gruber, A., & Campbell, T. L. (2015). *Companion robot for personal interaction*. Google Patents.

Weiss, A., Buchner, R., Tscheligi, M., & Fischer, H. (2011). Exploring human-robot cooperation possibilities for semiconductor manufacturing. *2011 International Conference on Collaboration Technologies and Systems (CTS)*, 173–177.

Wilson, G., Pereyda, C., Raghunath, N., de la Cruz, G., Goel, S., Nesaei, S., Minor, B., Schmitter-Edgecombe, M., Taylor, M. E., & Cook, D. J. (2019). Robot-enabled support of daily activities in smart home environments. *Cognitive Systems Research*, *54*, 258–272.

Wu, C.-G., Gerlach, J. H., & Young, C. E. (2007). An empirical analysis of open source software developers' motivations and continuance intentions. *Inf. Manage.*, *44*(3), 253–262.

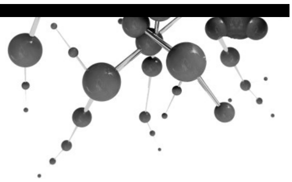
Zaman, B., Van Mechelen, M., & Bleumers, L. (2018). When toys come to life: Considering the internet of toys from an animistic design perspective. *Proceedings of the 17th ACM Conference on Interaction Design and Children*, 170–180.

Zanatto, D., Patacchiola, M., Goslin, J., Thill, S., & Cangelosi, A. (2020). Do humans imitate robots? An investigation of strategic social learning in human-robot interaction. *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 449–457.

Zhan, K., Zukerman, I., Moshtaghi, M., & Rees, G. (2016). Eliciting Users' Attitudes toward Smart Devices. *Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization*, 175–184.

Zhang, L., Jiang, M., Farid, D., & Hossain, M. A. (2013). Intelligent facial emotion recognition and semantic-based topic detection for a humanoid robot. *Expert Systems with Applications*, *40*(13), 5160–5168.

Zhuang, F., Zupan, C., Chao, Z., & Yanzheng, Z. (2008). A cable-tunnel inspecting robot for dangerous environment. *International Journal of Advanced Robotic Systems*, *5*(3), 32.



Paper I

Kertész, C., & Turunen, M. (2018). Exploratory Analysis of Sony AIBO Users. AI & Society, 1–14. doi:10.1007/s00146-018-0818-8

AI & SOCIETY (2019) 34:625–638
https://doi.org/10.1007/s00146-018-0818-8

OPEN FORUM

Exploratory analysis of Sony AIBO users

Csaba Kertész[1] · Markku Turunen[1]

Received: 11 October 2017 / Accepted: 23 January 2018 / Published online: 3 February 2018

Abstract
It is important to understand how the cultural background, the age and the gender influence the expectations towards social robots. Although past works studied the user adaptation for some months, the users with multiple years of ownership (heavy users) were not subjects of any experiment to compare these criteria over the years. This exploratory research examines the owners of the discontinued Sony AIBO because these robots have not been abandoned by some enthusiastic users and they are still resold on the secondhand market. 78 Sony AIBO owners were recruited online and their quantitative data were analyzed by four independent variables (age, gender, culture, and length of ownership), user contribution and model preference points of view. The results revealed the motives to own these robots for years and how the heavy users perceived their social robots after a long period in the robot acceptance phase.

Keywords Quantitative research · Heavy users · Social robot · Sony AIBO

1 Introduction

Nowadays, more and more social robots are introduced onto the market and the user expectations must be understood for the researchers to execute successful long-term experiments and for the companies to create sustainable business plans. Graaf et al. (2014) developed a theoretical foundation to describe the relationship between the owner and its robot over time. Before the purchase, the consumer learns about the technology and makes the decision to acquire the product (pre-adoption phase). In the adoption and adaptation phases, the first experiences are gained with the robot at home. When the novelty effect fades away and the user expectations are met, daily routines are developed with the robot (incorporation phase). After 6 months, the owner gets emotionally attached to the robot as a personal object (identification phase), the robot is finally accepted for long-term use and the owner becomes a heavy user.

The commercial Sony AIBO robots still have a reachable, significant user base long after the discontinuation. Although these users can be analyzed from many perspectives, this study was based on a questionnaire to measure the perception of the robot and the technical expectations. The authors concentrated on these aspects to identify the key needs of this community because the literature examined the phases before the robot acceptance, but we do not have a good understanding of the heavy users beyond the identification phase. This paper is a preliminary step to build this knowledge.

1.1 Robotics questionnaires in the literature

This section reviews past works whose subjects filled robotics questionnaires and they were analyzed by independent variables (culture, age or gender). These participants were not robot owners and they were recruited from the internet, at universities or exhibitions thus their preferences represented the general public to some extent.

Zhan et al. (2016) studied the attitudes towards smart devices with mainly Australian participants to develop a Recommender System for particular tasks. The male participants were more likely to accept robots than the female

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00146-018-0818-8) contains supplementary material, which is available to authorized users.

✉ Csaba Kertész
csaba.kertesz@ieee.org

Markku Turunen
markku.turunen@sis.uta.fi

and the age did not influence their ratings except the robotic pet which was disliked over 50 years of age.

Nomura et al. (2005) examined the visitors at an exhibition of interactive robots in Japan. Their questionnaire revealed that younger people liked the robots less than elder people and there was no difference in the behavior towards the robots from gender point of view. The visitors met with the robots once; therefore, these results were obtained in the pre-adoption phase.

Haring et al. (2014) developed a questionnaire to compare the European culture with Japanese regarding the emotions towards robots. It was found that Japanese people had higher exposure to the robots through the media, but they had less personal experiences than Europeans. This research did not found more positive attitude in the Japanese culture towards robots and Europeans accepted a human-like robot less than the Japanese people who saw the robots more as a machine.

Ezer (2008) explored in his PhD thesis what kind of roles the American public expects from a robot. His questionnaire was sent via mail to random individuals in Atlanta (USA) and their answers indicated age-related differences in the desired robot tasks which can be considered for robot design to achieve better utility value for different age groups. However, the survey did not show any pictures of actual robots to the participants and these results were limited by the general imagination about the robots that the media and the television programs show us.

Bartneck et al. (2007) studied members of online robotics communities and university students from seven countries. Their Negative Attitude towards Robots Scale (NARS) questionnaire was analyzed with three independent variables (community, gender, culture) and the female participants were more positive about the social influence of the robots than males. The cultural background had a significant influence on the initial attitude towards robots although Japanese people were not so positive as stereotypically assumed. Interacting with AIBO had a positive effect on the results, but owning a robot did not improve the acceptance.

A humanoid robot (Robi) was exposed to Japanese and Australian people in Haring et al. (2015) to find changes in the perception during an interaction. The first questionnaire filling before the session (Phase 0) reflected the cultural background and the initial expectations. The responses showed that the Australian participants liked the robot more and they rated its intelligence level higher than the Japanese people.

Nomura et al. (2012) measured the social acceptance towards humanoid robots in the Japanese population. They conducted a survey online with 1000 randomly selected persons to represent all age groups between 20 and 60 years. The older generations had more positive expectations what could be further improved with human–robot interactions (HRI) or robotics news in the media. After HRI experiences, the younger generations had increased apprehension towards robots and their anxiety did not pass.

None of the studies in this section reviewed actual users of a robot or the effects after extensive interactions thus all participants were before pre-adoption stage. If the questionnaire in this paper can reveal similar tendencies to (Bartneck et al. 2007; Nomura et al. 2005, 2012; Haring et al. 2014, 2015; Zhan et al. 2016) then the new results may generalize to the population.

1.2 Studies in human–robot interaction

The literature of HRI studied several robots in the past. A general observation was that the users have decreasing interest in the robots after their novelty fades away (Gockley et al. 2005; Salter et al. 2004). To avoid losing attraction, social or other engaging capabilities must be identified to create robots for longer use. The long-term interaction was studied by Leite et al. (2013) in a comprehensive survey of exploratory papers of health care, education, work and home settings. They admitted that the reviewed experiments were carried out with limited number of users and the purpose of the longer duration was to let the participants to get comfortable with the experimental conditions. Their results suggested that people were happy to interact with the social robots for longer periods, but they proposed further analysis to confirm this hypothesis.

Graaf et al. (2014) investigated the social robot acceptance in domestic environment. 70 Karotz robots, an internet-connected bunny-shaped ambient electronic device, were given to participants and their acceptance was tracked over 7 months. Many participants lost interest after the initial excitement disappeared, but 10% of the subjects remained active until the end of the experiment.

In Fernaeus et al. (2010), a dinosaur robot (Pleo) was given to families for several months and it was studied how the experiences met with the high user expectations due to the price and the advertisements. In the reality, the robot skills were not enough sophisticated to satisfy the participants after the initial novelty effect faded away and Pleo was switched on rarely.

This study focuses on a domestic social robot, namely, the owners of Sony AIBO robots (Fig. 1) were analyzed. This product brand included quadruped autonomous entertainment robots which had a behavior-based architecture to exhibit a life-like impression. These robots walk around the room, interact with the owner and switch between probabilistic state machines to show rich behaviors and engage the owners. Several papers have been focused on AIBO in the past decade, but the heavy users were not examined. In Friedman et al. (2003), the online forums for Sony AIBO were analyzed to investigate the relationship between the robots and their owners. According to

Fig. 1 Sony AIBO robots

the posts on the internet, people developed an emotional connection to these robots, but they rarely attributed moral standing towards them.

The temporal change of an attitude can be examined with longitudinal studies where the same people are tracked over months (Coninx et al. 2016; Fernaeus et al. 2010; François et al. 2009; Graaf et al. 2014), but this is time-consuming with unexpected technological complications and scheduling user interviews. To the best knowledge of the authors, all past HRI researches organized short weekly or monthly sessions for the participants together with the robot (Coninx et al. 2016; Fernaeus et al. 2010; François et al. 2009; Koay et al. 2007) before reaching the acceptance phase. The participants in this paper owned Sony AIBO for years and they still run these robot dogs time to time what is a fundamental difference from, e.g., Fernaeus et al. (2010). The earlier studies had also a challenge to recruit enough participants to allow statistical calculations for significant trends in the data, except (Graaf et al. 2014).

The subjects in this experiment lived with commercial, social robots day by day for years. Since Sony AIBO was discontinued long ago when this survey was conducted, people were reachable who owned these robots for more than 10 years. Furthermore, the Sony AIBO community on the internet was still active with new members who bought these social robots from secondhand sources. At the same time, newcomers and experienced owners could participate in this experiment; therefore, the subjects of this paper were beyond the acceptance phase and used their robots for years. The paper tries to answer the following research questions:

- How does the length of ownership affect the perception of the robot?
- Is there any significant difference between Westerners and Japanese people?
- Does the age (young/middle age/old) change the users' opinion?
- Does the gender make any difference?

2 Questionnaire

A questionnaire (Appendix) was conducted to get the opinion of people about their Sony AIBO robots. A flexible design was chosen with many Likert-type items which were easy to understand and fill out, but the participants also had the chance to give their own opinion in optional text fields. The expectation was that the target group (heavy users) had constant interactions with their robots; therefore, one question ensured that the participants run their social robot regularly. The other questions were related to the perception of their robots, how they feel about the existing software and which skills must be improved in AIBO. Eight questions asked basic information about the participants (gender, age, home location, profession) and the robot ownership (length, usage frequency, model preference). Question 9 investigated the impressions about the existing AIBO skills with 9-point Likert-type items (anchors: 1—strongly disagree, 3—disagree, 5—neutral, 7—agree, 9—strongly agree) and the following two concentrated on the expectations from a new software update. The adjectives of the Likert-type items for emotional perception were selected together with a psychologist who worked in the social robotics field and had experiences with questionnaires. Other Likert-type items queried technical aspects to identify wishes for specific skills of these social robots and the remaining questions collected answers about the connectivity options, autonomous behavior and possible user contributions. Several questions included text fields where the participants could enter additional comments in a free form, but those answers were analyzed in a previous paper of the authors Kertész and Turunen (2017).

78 fillings were collected from the members of an English speaking online AIBO forum (http://aibo-life.org) what is similar to 70 in Graaf et al. (2014) and 17 Japanese participants were reached via Facebook ad campaign, similar to Samuels and Zucco (2012). Although Bartneck et al. (2007) distinguished the culture based on the nationalities, the current survey clusters the 61 forum members to Western culture and the 17 Japanese responses to Japanese culture. The authors made this decision to examine the stereotypical belief, similar to Bartneck et al. (2007), that the Japanese people consider the robots with soul, unlike in the Western culture where the robots are recognized more as machines. Sony AIBO was a major hit in the Japanese market and the remaining user base was significant to form a single group in this survey for the culture variable.

The questionnaire was filled by 57 males and 19 females, since two participants did not reveal their gender, with a ratio 73%/24%, similar to the reported 69%/31%

gender ratio of AIBO owners in Fujita (2004) and another online AIBO questionnaire with 64%/36% in Bartneck et al. (2007). Although there was no question about the income and the wealth of the participants, the authors can explain this rate with the possible higher interest of the men in gadgets and they can afford more to buy expensive robots (Yoldas 2012; Zhan et al. 2016). The technical enthusiasm was also reflected in the professions because most owners were engineers, software developers or technicians (27% for Tech in Fig. 2b) and other occupations were between 1–15% in Fig. 2b.

Since this study focused on robot consumers with more years of ownership after technology acceptance phase, the analysis does not include such people who did not go beyond the technology adoption with Sony AIBO. Almost the half of these owners was young adults (under 40 years) or in middle age (40–60 years) and 5.13% was old (Fig. 3a). It is worth noting that 14.10% of the participant was less than 25 years and their stories on the online forum given an insight of their intentions to buy these robots. Either they have got know AIBO in the recent years or they were children during the commercialization of AIBO and they could afford these robots after becoming an adult with income. The measured age distribution is very similar to the statistics of Japanese Sony AIBO owners reported in 2002 (Fujita 2004) what was an interesting similarity between the commercial and secondhand periods in the product lifecycle. Figure 3b shows high retention rate for 20.51% of the participants who kept their robots for more than 10 years, 51.29% had AIBO between 2 and 10 years, but 28.21% possessed AIBO for less than 2 years which is a high rate of newcomers.

Two questions in the survey asked about the age and the length of ownership. The first age group was under 25 years (age$_1$), the second between 25 and 40 years (age$_2$) and the third over 40 years (age$_3$). The length of ownership had four ranges: less than 2 years (years$_1$), 2–5 years (years$_2$), 5–10 years (years$_3$) and over 10 years (years$_4$). The majority of young generation (age$_1$) possessed their robots for less than 2 years (years$_1$) on Fig. 4, the most typical duration for age$_2$ group was between 5 and 10 years (years$_3$) and the biggest portion of the oldest age group (age$_3$) owned their gadgets over 10 years (years$_4$). These results suggest that elder generations keep their robots longer and adults over 25 years (age$_2$ and age$_3$) had their Sony AIBO for varying years (years$_1$–years$_4$), the latter generations are newcomers and long-term customers at the same time.

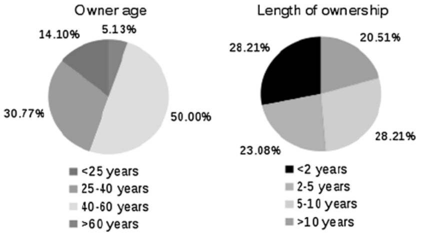


Fig. 3 The age (**a**) and length of ownership (**b**) of the participants

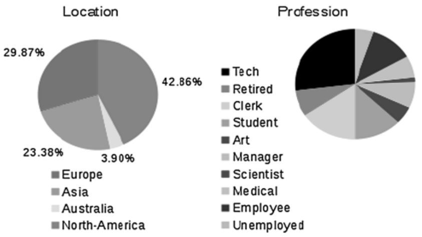


Fig. 2 The home location (**a**) and profession (**b**) of the AIBO customers who responded to the questionnaire

3 Overall analysis

Most questions in the survey were constructed with Likert-type items, they were grouped into subscales and their consistency was analyzed with Cronbach's α coefficients for sufficient trust in the overall reliability:

1. Emotional perception of the robot.
2. Emotional expectations from a new software.
3. Expected skill improvements in a software update.
4. Connectivity options to the robot.

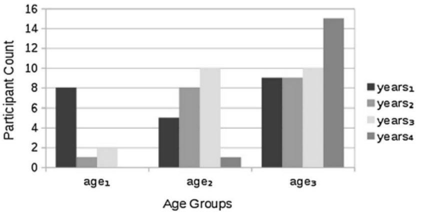


Fig. 4 Age distribution (age$_1$ = < 25 years, age$_2$ = 25–40 years, age$_3$ = > 40 years) of the questionnaire participants in the function of the length of ownership (years$_1$ = < 2 years, years$_2$ = 2–5 years, years$_3$ = 5–10 years, years$_4$ = > 10 years)

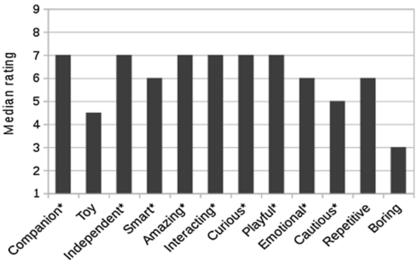


Fig. 5 Median ratings of the existing robot software for Sony AIBO robots. The items in the second factor of the exploratory factor analysis are marked with an asterisk

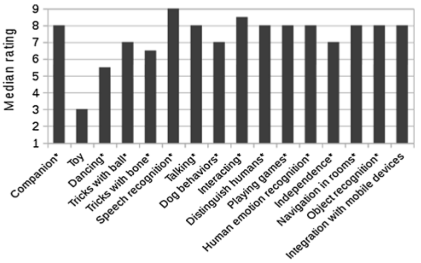


Fig. 7 Median ratings for improvements in the current robot software of Sony AIBO. The items in the first factor of the exploratory factor analysis are marked with an asterisk

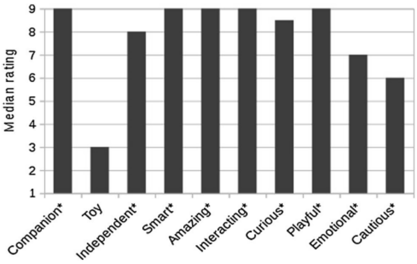


Fig. 6 Median ratings for wished features in a new software of Sony AIBO. The items in the fourth factor of the exploratory factor analysis are marked with an asterisk

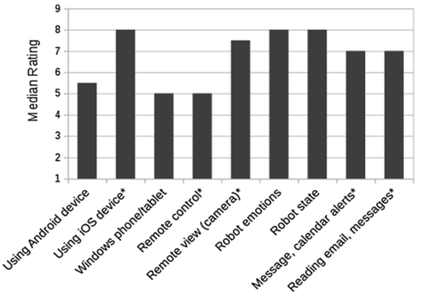


Fig. 8 Median ratings for wished connection options in a new software of Sony AIBO. The items in the third factor of the exploratory factor analysis are marked with an asterisk

The questionnaire did not have many responders (78) for an ideal quantitative data analysis, but the subscales had good α coefficients (0.82, 0.87, 0.91, 0.81).

3.1 Exploratory factor analysis

Exploratory factor analysis is a statistical method for uncovering the underlying structure of a questionnaire. All Likert-type items were investigated by this method to verify if the answers were coherent inside the subscales.

The Kaiser–Meyer–Olkin Measure of Sampling Adequacy was 1, the Bartlett's Test of Sphericity was significant under $p < .001$ for approximate of Chi-Square 3649.37 thus the measured variables were not normally distributed, but skewed. 12 items had eigenvalue over 1.00 and they expressed 78.17% of the total variance. Further evidence was for the coherence that several factors included most questions of certain subscales. Namely, the first subscale was found in the second factor (Fig. 5), the second subscale in the fourth factor (Fig. 6), the third subscale in the first factor (Fig. 7) and the fourth subscale in the third factor (Fig. 8).

The correlated items in the first factor were from the third subscale where AIBO owners wished to receive software updates to improve the robot intelligence (Fig. 7) except the toy skill. The second factor (Fig. 5) were from the first subscale and the owners perceived their robots with a positive attitude. They treated these machines as independent companions who were smart, amazing, interacting, curious, playful and emotional.

Some factors indicated relationship between items on a lower granularity:

- In the third factor, the Apple customers wished remote camera viewing, reading and alerts of messages and calendar events. On the other hand, Android and Windows

Phone users preferred to control the robot remotely (fifth factor).

- People would not find more autonomous software for the robot disappointing or boring according to the fourth factor.
- The sixth factor showed that the participants did not regard AIBO as a toy and they would not like to see improvements in features to make the robot a toy.
- Those owners would feel curious and fun a more autonomous mode in new software who finds the existing personalities boring or repetitive according to the seventh factor.

3.2 Subscale results

Although Sony AIBO is an entertainment robot with limited capabilities, people attributed positive, life-like properties (e.g., companion, independent, smart) to the robot with 5–7 median ratings (Fig. 5) in the first subscale and it was not found boring (3) or a toy (4.5). On the other hand, the original software of AIBO was found quite repetitive (6) and the owners desired more sophisticated social skills from new software with median ratings (7–9) of the second subscale on Fig. 6. Thus, the owners judged these robots really "social" instead of a toy after years. This outcome was an opposite of an earlier experiment with Pleo in which the participants already treated the dinosaur robot a toy after some months (Fernaeus et al. 2010).

The third subscale in the questionnaire focused on particular feature updates in the current software (Fig. 7). Enhancing the dances (5.5) and toy-like functions (3) had again lower interest. The tricks with plastic toys of the robot (ball, bone) and dog-like behaviors were not found so vital topics, most likely, because these features provide the entertainment aspect of the current software and the people are more eager to interact with the robots. The human–robot interaction skills had the highest median ratings: companion (8), speech recognition (9), talking (8), interacting (8.5), distinguish humans (8), playing games (8), and emotion recognition (8). These skills shape a valuable emotional connection between the robot and people instead of watching repetitive entertainment behaviors. Worth to note that the participants would have liked to have enhanced autonomous features (navigation in rooms, object recognition) and further connection options following up the trend of portable handsets and tablets (integration with mobile devices) in the recent years. This result was aligned with Fig. 9 where the median ratings suggested that the robot customers had positive anticipations about a more autonomous software for Sony AIBO (Fig. 9).

The connectivity options were queried in the fourth subscale (Fig. 8) and the majority was grouped in the third factor. The robots and Apple products are expensive, therefore, the higher classes can afford these items with a higher chance than people with lower income and they could be overrepresented in the survey responders although there was no question related to their wealth. The participants wished to connect their bots to iOS devices with the highest ranking (8) while Android had a moderate result (5.5), and according to the low market share, Windows devices had the lowest rating (5). Interestingly, people would like to see the robot state, emotions and camera image with median rankings 7.5–8, but the remote control had a lower interest (5). This phenomenon can be originated that the people can associate a remote controlled robot to a soulless machine and making AIBO a toy was not an unattractive skill for the robot owners in Figs. 5, 6 and 7. Receiving alerts about SMS, calendar events or reading email were popular and scored to 7.

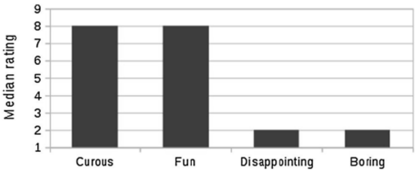


Fig. 9 Median ratings of the anticipated feelings caused by an autonomous personality in a forthcoming software for Sony AIBO

To sum up this section, the exploratory factor analysis confirmed that the subscales were defined consistently and reliable results can be expected from the further examination. Looking at the median ratings of the Likert-type items, the participants had great desire to make the robot an autonomous companion which can interact with humans and connect the robot with the latest gadgets, but the repetitive behaviors for entertainment and acting as a toy were out of interest.

4 Analysis of the independent variables

The questionnaire responses were examined from gender, age, length of ownership and culture points of view to see how these independent variables affected the ratings. Each variable had a corresponding null hypothesis to be examined. If a variable had two categories, the Likert-type items were evaluated with Mann–Whitney test, for more than two categories, Kruskal–Wallis tests were performed. Depending on the results of these tests, the null hypotheses were either accepted or rejected.

Any difference in the median ratings of the Likert-type items was reported in the following sections when they were higher than 1 and their *p* values were less than 0.25. In this way, the paper includes some insignificant results, but the authors wanted to keep these items because they fit in the tendencies of the significant items. Greater *p* values were ignored.

In the last two sections, the user contribution and the robot model preference were analyzed by the independent variables separately because these questions did not use Likert-type items.

4.1 Gender

Null hypothesis (H1) The male heavy users see a social robot as a machine and the female as a companion.

57 participants were male and 19 female from the survey respondents. The gender defined an independent variable with two categories and all Likert-type items were evaluated with Mann–Whitney test. The median ratings of women were all the time higher for those items where the robot was treated as a living being, the significant differences are shown on Fig. 10. Women attributed more human feelings to the existing software, hence (Bartneck et al. 2007) companion ($p = .024$), (Bartneck et al. 2007) smart ($p = .004$), (Bartneck et al. 2007) amazing ($p = .021$) and (Bartneck et al. 2007) curious ($p = .000$) received more positive ratings by 2. This attitude was also reflected in their wishes for improvements with (Fernaeus et al. 2010) Human emotion recognition ($p = .085$).

Males rated (Bartneck et al. 2007) toy ($p = .004$) and (Coninx et al. 2016) toy ($p = .004$) with higher median by 2, therefore, they regarded the robot to a greater extent as a machine. Similarly, men found more important to read e-mails and messages by the robot because they rated (François et al. 2009) reading email, messages ($p = .031$) higher by 2.

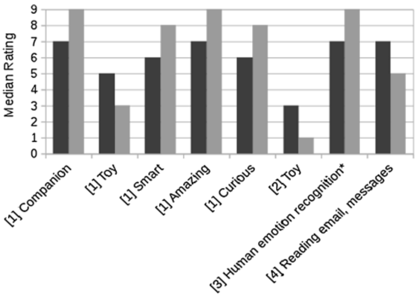


Fig. 10 Median ratings for items with considerable difference between male (blue) and female (red) answers whose number in square brackets refers to the subscale of each item. All components have significance under 0.05 except those which are marked with an asterisk

As the common sense suggests, women tended to be more emotional in their ratings while the men were technology-minded. The H1 null hypothesis was accepted for the gender variable.

4.2 Age

Null hypothesis (H2) The younger heavy users are more technology-minded while the elder look the social robots as a companion.

11 participants were under 25 years (age$_1$), 24 between 25 and 40 years (age$_2$) and 43 over 40 years (age$_3$). The age defined an independent variable with three categories and all Likert-type items were evaluated with Kruskal–Wallis test. Figure 11 shows the items from the first subscale which had a significance value below or close to 0.050 and two main tendencies can be observed. On one hand, the older people the less positive they were about scoring the existing skills of the robot what can be seen for (Bartneck et al. 2007) interacting ($p = .047$), (Bartneck et al. 2007) curious ($p = .027$) and (Bartneck et al. 2007) emotional ($p = .000$) on Fig. 11. On the other hand, the older age groups associated the robot with more negative properties by (Bartneck et al. 2007) toy ($p = .096$), (Bartneck et al. 2007) repetitive ($p = .065$) and (Bartneck et al. 2007) Boring ($p = .008$). The exceptional (Bartneck et al. 2007) Cautious ($p = .011$) age$_2$ found the robot more cautious than age$_1$ and age$_3$ groups.

There was no significant difference between the answers in the second subscale, but the third and fourth subscales revealed on Fig. 12 that the age$_2$ group tended to be more eager to see improvements in autonomous features of a new software with (Fernaeus et al. 2010) tricks with ball ($p = .015$), (Fernaeus et al. 2010) tricks with bone ($p = .013$)

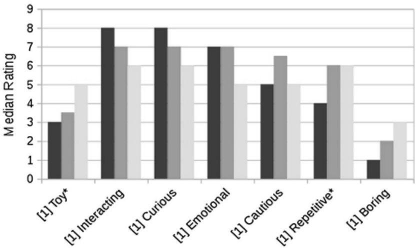


Fig. 11 Median ratings for items in the first subscale between age groups (blue for age$_1$, red for age$_2$ and yellow for age$_3$). The number in square brackets refers to the subscale of each item. All components have significance under 0.05 except those which are marked with an asterisk

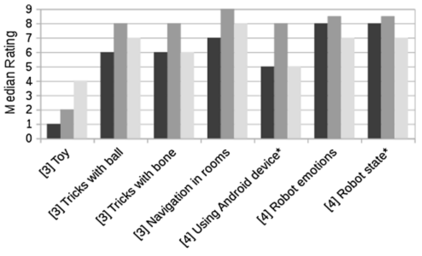


Fig. 12 Median ratings for items in the third and fourth subscales between age groups (blue for age$_1$, red for age$_2$ and yellow for age$_3$). The number in square brackets refers to the subscale of each item. All components have significance under 0.05 except those which are marked with an asterisk

and (Fernaeus et al. 2010) navigation in rooms ($p = .013$). Similar to this trend, age$_2$ was most interested in connectivity with Android phones and view (François et al. 2009) Robot emotions ($p = .011$) although (François et al. 2009) using Android device ($p = .059$) and (François et al. 2009) Robot state ($p = .113$) were not significant. Older age groups recognized the robot rather as a toy (Fig. 11) and the same trend was measured for the wish of improved toy features with (Fernaeus et al. 2010) toy ($p = .027$) in Fig. 12. This may be originated in some reappearing traits from childhood in old age.

The results did not reflect the expectations of the null hypothesis (H2), elder people did not perceive these social robots as a companion to a greater extent and younger generations were not more eager about the technology side of these robots. Therefore, H2 was rejected.

4.3 Culture

Null hypothesis (H3) The Japanese heavy users do not rate their robots more positively than Westerners.

The cultural background of the participants was examined to compare Western people with Japanese. 61 responders were from Europe, North-America and Australia. Since these samples were collected from an English speaking AIBO forum and a closed Facebook group, the authors made sure that it did not contain any responses from migrated Japanese citizens in these countries. This original version of the questionnaire was in English.

17 Japanese fillings were gathered with a targeted Facebook ad campaign in Japan. The survey was localized to Japanese language and the filters of the campaign ensured that native Japanese people filled this variant. Additional evidences of the correct sampling were a special answering pattern to the occupation question by Japanese people and the common lake of answers to the free-form entries.

The culture defined an independent variable with two categories; therefore, all Likert-type items were evaluated with Mann–Whitney test. The existing programs and wishes for new software were slightly less attractive for the Japanese in Fig. 13. The median ratings of (Bartneck et al. 2007) Interacting ($p = .001$), (Bartneck et al. 2007) curious ($p = .017$), (Bartneck et al. 2007) playful ($p = .006$), (Bartneck et al. 2007) emotional ($p = .008$), (Coninx et al. 2016) companion ($p = .047$), (Coninx et al. 2016) amazing ($p = .090$), (Coninx et al. 2016) interacting ($p = .001$) and (Coninx et al. 2016) playful ($p = .150$) followed tendencies of the Westerners, but they were lower by 2. Figure 14 has a similar pattern, the Japanese participants scored the feature

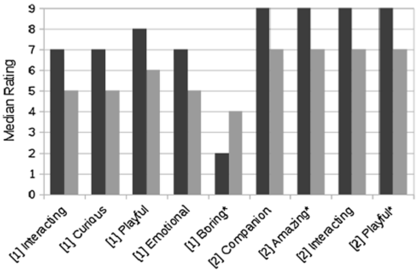


Fig. 13 Comparison of median ratings in the first and second subscales with different cultural backgrounds (blue for Westerners, red for Japanese). The number in square brackets refers to the subscale of each item. All components have significance under 0.05 except those which are marked with an asterisk

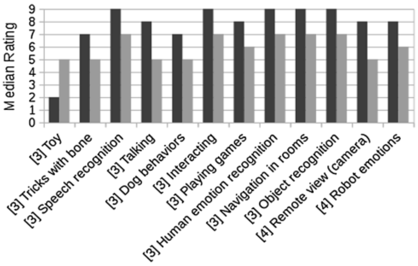


Fig. 14 Comparison of median ratings with different cultural backgrounds (blue for Westerners, red for Japanese). The number in square brackets refers to the subscale of each item. All components have significance under 0.05 except those which are marked with an asterisk

improvements and connectivity options lower by 2–3 with (Fernaeus et al. 2010) tricks with bone ($p = .011$), (Fernaeus et al. 2010) speech recognition ($p = .008$), (Fernaeus et al. 2010) talking ($p = .006$), (Fernaeus et al. 2010) dog behaviors ($p = .005$), (Fernaeus et al. 2010) interacting ($p = .033$), (Fernaeus et al. 2010) playing games ($p = .017$), (Fernaeus et al. 2010) human emotion recognition ($p = .008$), (Fernaeus et al. 2010) navigation in rooms ($p = .047$), (Fernaeus et al. 2010) object recognition ($p = .002$), (François et al. 2009) remote view ($p = .001$) and (François et al. 2009) robot emotions ($p = .001$). Furthermore, Japanese found the existing software quite (Bartneck et al. 2007) boring ($p = .081$) in Fig. 13, but they wanted to see enhancements in Fernaeus et al. (2010) Toy ($p = .004$) features in Fig. 14.

Since the Japanese people were more negative about Sony AIBO, the null hypothesis (H3) was accepted.

4.4 Length of ownership

Null hypothesis (H4) The more years a heavy user owns a social robot without content updates the more robot acceptance decreases and he/she loses interest over time.

The questionnaire results were analyzed in the function of length of ownership with four categories. 22 participants have been owned the robots for less than 2 years (years$_1$) and 18 between 2 and 5 years (years$_2$). These people were new members in the AIBO community years after the product discontinuation. 22 have been experienced these robots between 5 and 10 years (years$_3$) and 16 over 10 years (years$_4$). Because the independent variable had more than two categories, the Likert-type items were evaluated with Kruskal–Wallis test.

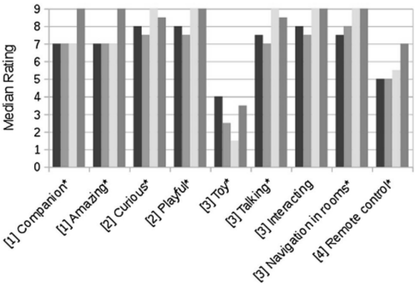


Fig. 15 Comparison of median ratings of participants with different length of ownerships (blue for years$_1$, red for years$_2$, yellow for years$_3$ and yellow for years$_4$). The number in square brackets refers to the subscale of each item. All components were marked with an asterisk because their significance were over 0.050

Figure 15 shows nine items and almost all items had significance value over 0.050. Despite the common sense suggests that the consumer interest must decline after years of usage without software updates, the results did not reflect this expectation. In particular, the existing software was evaluated in the first subscale and there were no significant decline in the anthropomorphic characterization. The owners with more than 10-year-long experience rated the robot with existing software more (Bartneck et al. 2007) amazing ($p = .137$) and a (Bartneck et al. 2007) companion ($p = .266$) while they wished more (François et al. 2009) remote control ($p = .190$). After 5 years of ownership, the need for autonomous and social features was increased. Years$_3$ and years$_4$ groups were wished a more (Coninx et al. 2016) curious ($p = .154$) and (Coninx et al. 2016) playful ($p = .108$) robot which is eager for (Fernaeus et al. 2010) talking ($p = .165$) and (Fernaeus et al. 2010) interacting ($p = .011$) with its owner. (Fernaeus et al. 2010) toy ($p = .160$) feature was an odd-one-out feature because longer ownership decreased the desire for a toy robot although the years$_4$ group broke this tendency. In an earlier work (Bartneck et al. 2007), the ownership did not influence the attitude of the people towards robots, but Fig. 15 suggests that the owners will appreciate their robots more after 5 years. The null hypothesis (H4) was rejected because of these results.

4.5 User contribution

One question asked the owners if they would contribute new tricks for the robot software with a motion editor application. Figure 16 shows that almost two-third of the users (58.97%) expressed willingness to create new content for the robot, one-third (30.77%) refused and 10.26% was unsure. As Sect. 4.1 confirmed, men were more focused on the technology side of the robot and males expressed double chance (64.42%) for contribution compared to females (31.58%) in Fig. 16 and women were twice as much unsure (males 7.02% vs. females 15.79%).

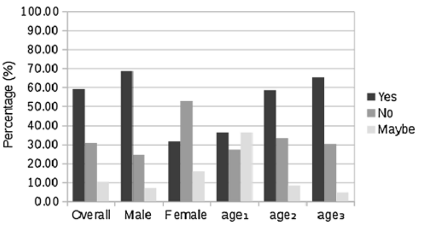


Fig. 16 Overall answers for user contribution question and details for gender and age (age$_1$ = < 25 years, age$_2$ = 25–40 years, age$_3$ = > 40 years) conditions

The traditional stereotype suggests that elder people are less open to the new technologies and younger generations catch them up more easily. Figure 16 presents the opposite, the elder the owners are, the higher chance they will use their computer to create new motion content (age$_1$: 36.36% < age$_2$: 58.33% < age$_3$: 65.12%) and older generations were less hesitant about to make this decision (age$_1$: 36.36% > age$_2$: 8.33% > age$_3$: 4.65%). Otherwise, all age groups rejected the contribution around 30%.

When the answers were analyzed from the length of ownership condition, the people with over 10-year-long ownership (years$_4$) were the most eager (87.50%) to create new motions for their robots, other owners given around 50% (years$_{1–3}$). It is noteworthy that the newcomers (years$_1$) were the most hesitant (22.73%) to make any content; the half of the owners with 2–10-year-long experience (years$_{2–3}$) had no intention for contribution.

Japanese society is regarded acceptant for high-tech compared to other Western countries. This phenomenon is reflected in the culture condition on Fig. 17. 82.35% of the Japanese owners were keen to create new motions for the robot, 17.65% were not interested in and none of them were unsure.

4.6 AIBO model preference

The Sony AIBO brand had several generations and every model had different advantages and disadvantages. The following list describes the most important characteristics of each product without going into technical details. These points were considered by the buyers when they decided to acquire a specific model:

- ERS-1xx: The first model in the series. More autonomous and advanced emotional personality was implemented in these robots. It had a charging station, but it could not execute the self-docking process. Despite the fragile body or repetitive behaviors many owners liked ERS-1xx because of the high degree of freedom, intensive interactions in the growth stages of the personality and the impulsive exploration mode. Secondhand price: $300–400.
- ERS-2xx: The emotional model was not so complex in this successor. Many softwares were available for ERS-2xx, but each provided different features (e.g., self-docking, autonomous exploration). The owner had to change the memory card to switch between the personalities. The appearance of these models was much more clean compared to ERS-1xx. Secondhand price: $500–700.
- ERS-3xx: This model was a simplified version of ERS-2xx with a lower price tag and the cute design to target female customers (Fujita 2004). It had no wireless connectivity and less software was shipped for ERS-3xx. Secondhand price: $400–500.
- ERS-7: The latest model had the best hardware and all previous skills of ERS-2xx were combined into a single firmware (Mind software). The robot docked itself to the charging station, did tricks, communicated with the owner, played with its toys and explored the surroundings although the latter skill was not as adventurous as in ERS-1xx. Secondhand price: $1800–2200.

The most advanced model was the Sony ERS-7 which still maintains high price on the secondhand market. The ERS-2xx provided the same features scattered in various softwares and the lower price range ($500–700) balance this disadvantage. In general, ERS-7 and ERS-2xx were the favorite models in Figs. 18 and 19 and ERS-7 was the top rated for being the most advanced. Third place went for the ERS-1xx whose interaction skills and autonomous mode compensated the weak hardware design and the lack of features. The ERS-3xx models were rated worst caused by the missing connectivity and software options. When the user preferences were reviewed for the gender condition, the

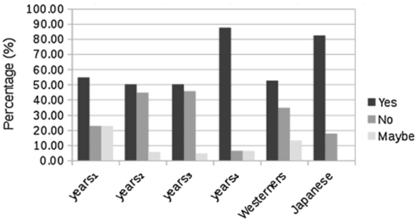


Fig. 17 Detailed answers for the user contribution question from length of ownership (years$_1$ = < 2 years, years$_2$ = 2–5 years, years$_3$ = 5–10 years, years$_4$ = > 10 years) and culture conditions points of view

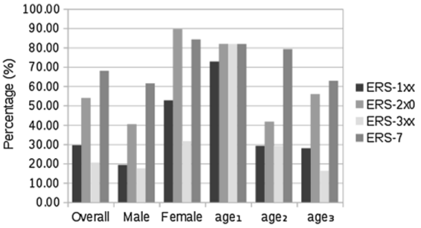


Fig. 18 Preferred AIBO models. One participant could select multiple favorite models. This diagram contains the overall results and the details for gender and age (age$_1$ = < 25 years, age$_2$ = 25–40 years, age$_3$ = > 40 years) conditions

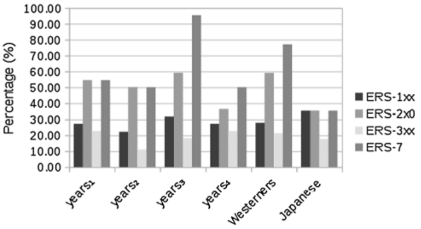


Fig. 19 Preferred AIBO models. One participant could select multiple favorite models. This diagram contains detailed answers from length of ownership (years$_1$ = < 2 years, years$_2$ = 2–5 years, years$_3$ = 5–10 years, years$_4$ = > 10 years) and culture conditions points of view

males chosen a favorite model less frequently than females and the women preferred ERS-2 × 0 in the first place over ERS-7 exceptionally.

The age condition revealed in Fig. 18 that young people (age$_1$) had no clear preference for any Sony AIBO, they were interested in all models. The ERS-2xx and ERS-7 series were in tie for < 10 years in the length of ownership condition (Fig. 19), but the latest model was a winner for years$_3$ group. The Japanese people tend to select only one favorite which can be originated in the emotional attachment after buying their first model. The model preference of Westerners followed the overall results in Fig. 18 while Japanese rated all models to 35.29% except the ERS-3xx to 17.65%.

5 Discussion

5.1 Gender

The H1 null hypothesis was constructed with considering the differences between male and female social behavior and their social roles in society (Norman 2004) and it was accepted for the gender variable because the female participants were more emotional in their ratings while the men were more technology-minded. The past researches (Nomura et al. 2005, 2006, 2012) had mixed results within the Japanese society. A survey of female and male Japanese at an exhibition did not find a variance in their opinions about robots after an interaction (Nomura et al. 2005) and an online survey among Japanese people had the same result in Nomura et al. (2012). However, Japanese students were asked about the attitudes towards robots without actual interaction in Nomura et al. (2006) and the female students were more positive about the emotional interactions with robots than male, similar to H1. Most likely, there is no gender difference in the Japanese society according to Nomura et al. (2005, 2012), but some subgroups (younger generations?) can show some variation. On the other hand, female Italians attributed more positive feelings for robots than males in a questionnaire (Scopelliti et al. 2005) what strengthens H1 since the majority of the participants in this paper were from Western societies. The results in this regard were more close to cultural expectation according to the literature, similar to the next section.

5.2 Age

H2 was rejected because elder people did not perceive Sony AIBO as a companion and younger generations were not more interested in the technology side of these robots. This result is similar to Zhan et al. (2016) in which Australian participants over 50 years disliked the robotic pets, younger Italian generations had more positive feelings towards robots in Scopelliti et al. (2005) and Ezer reached the same conclusion in Ezer (2008) with American citizens, the acceptance was higher among younger adults in pre-adoption phase (without firsthand experience with robots). However, two studies (Nomura et al. 2005, 2012) found the opposite in the Japanese society. Younger Japanese generations liked the robots less than elder people, but this discrepancy can be explained with the different sampling. Nomura et al. (2005, 2012) were interviewed only Japanese while this paper included mainly Westerners. The revealed attitudes here and in the literature strongly suggest that the generations in the Japanese and Western cultures have the opposite preferences. However, the limited sample size was not enough to carry out more analysis about this assumption within this work.

5.3 Culture

The common stereotype suggested in the past that the Japanese people love robots more than Westerners. Some early attempts tried to understand and explain this belief by comparing the cultures (Kaplan 2004) instead of executing experiments with humans. Later, surveys (Bartneck et al. 2007; Haring et al. 2014) revealed the opposite, Japanese are not so positive towards robots and they are worried about the social and emotional impacts of robots in their society. The null hypothesis (H3) in this paper was constructed in accordance to these discoveries and the results confirmed H3, therefore, it was accepted. Worth to note that Japanese were more negative than Westerners in all cases what strengthens the results in (Haring et al. 2015) where Australian people were more positive about likeability and intelligence of a humanoid robot. Other works (Bartneck et al. 2007; Haring et al. 2014) showed no difference, but the results in this paper and (Haring et al. 2015) require further analysis to clarify the roots of the unexpected antagonism of the Japanese people towards robots.

5.4 Length of ownership

The rejection of H4 was a positive finding. The common sense suggests that people leave things behind after their utility value decreases over time, but the heavy users of Sony AIBO acted in the opposite way after 5 years of the ownership.

Getting people into acceptance phase is challenging because the high price tag of the robots induces high expectations from the users. Some participants did not finish a 6-month-long experiment with Roomba vacuuming robots (Sung et al. 2010) as they required too much maintenance. Though Graaf et al. (2014) analyzed the robot owners on a different timescale (half year vs. multiple years here) and 5–10% of their participants showed constant interest for Karotz robot in a longer period. Other people did not become a power user because Karotz robot did not offer more functions than a modern smart phone and they did not feel the robot useful, similar to other past works (Fernaeus et al. 2010; Leite et al. 2013).

The authors believe that the acceptance for Sony AIBO robots did not decrease over the years in the current study because their software was designed to develop emotional attachment with their owners and their main functions were not intended to replace a computer.

5.5 User contribution

The user contributions strengthened the results of Sect. 4.1 and older generations were more likely to contribute, especially after owning a robot for more than 10 years. It was also revealed that the Japanese participants were more eager and less hesitant to make technical contributions unlike Westerners. This finding is exciting compared to the long decision making in Japanese companies or the hesitation of Japanese people to say black-and-white decisions straightforward (Bernhauerova 2013). Maybe the anonymous questionnaire on the web allowed the Japanese participants to leave their comfort zone and they could really express what they think. The higher willingness for contribution of Japanese people was a bit surprising when this result was compared to the negative trends of Japanese answers for the culture variable (Sect. 4.3).

5.6 AIBO model preference

ERS-3xx had the worst score in almost all cases in Figs. 18 and 19 which suggests that people would like to use a social robot with rich skills and connectivity options. The low price tag cannot compensate the missing capabilities if similar robots with more advanced intelligence are on the market. Despite a Sony ERS-7 costs 3–4 times more than other series, people tend to prefer this model as a result of the most developed hardware and software.

5.7 Social robot design

Two papers proposed guidelines to improve the design of social robots for long-term human–robot interactions. Leite et al. (2013) presented a good review of this problem with a detailed discussion by accumulating the experiences of different robots in the research literature while (Graaf et al. 2016) expressed their recommendations on a higher level. The authors reviewed the free-form answers of the questionnaire in an earlier work (Kertész and Turunen 2017) and some proposals were already given to the literature. Based upon the quantitative analysis in this paper, additional suggestions are presented here in descending priority to complement the past works (Graaf et al. 2016; Leite et al. 2013):

- The long-term ownership does not bias the tendencies of robot acceptance what is expected in a certain cultural background and life stage. However, the literature review suggests that the acceptance must be examined with culture and age variables with sufficient sampling at the same time because the people's preferences vary in these dimensions significantly and they make hard to draw general conclusions.
- The age distribution of the owners remains stable, even long after the sales are stopped and the robot is traded on the secondhand market (see Sect. 2).
- The long-term ownership does not degrade the appreciation towards the robot after several years (see Sect. 5.3), but the owners desire the integration of the latest technologies.
- The robot should not replace the functions of a smart phone or a computer, especially with more hassle. The robot needs to differentiate itself from other machines with unique skills (see Sect. 5.3).
- Do not sacrifice essential skills to reduce the hardware costs otherwise consumers will not like the robot (see Sect. 5.6).
- The robot must adapt its personality with subtle differences according to the human gender, age and culture (Sect. 5.1, 5.2 and 5.4).
- Older people tend to treat the robot as a toy and they are less positive about the social skills, but they are more enthusiastic to create new content for the robot (Figs. 11, 12, 16).

5.8 Limitations

Despite of the participants were recruited on a special internet forum and Facebook on the internet, only 78 active Sony AIBO owners were reached, but the authors believe that sample size was reasonable compared to 230 in (Bartneck et al. 2007) and 41 in (Haring et al. 2014) considering that conducting our survey was long after the discontinuation of Sony AIBO. The sampling was not representative for the general public, but the participants could provide a good indication about the typical users of entertainment robots and even beyond this group since Bartneck et al. also found in (2007) that owning a Sony AIBO did not result significantly different scores on their NARS questionnaire.

Since these robots were commercial, this study was essential to analyze the heavy users of an expensive robot from the market. The robots in past experiments were given to participants on a voluntary basis for free (Fernaeus et al. 2010; Graaf et al. 2016).

The sample size was moderate, the cultural variable could not be evaluated on a nationality level, and therefore, this category was reduced to a comparison between Westerners and Japanese people. Unfortunately, the results seemed to be biased by the Western majority for the gender and age variables (Sects. 5.1, 5.2) when the results were compared to the literature. The authors still believe that the previous sections presented significant results and further exploratory studies are required with other social robots to confirm or complement the findings.

6 Conclusions

The heavy users of Sony robot dogs were studied in this paper after 10 years of the product discontinuation. Since these people owned their robots for years after the initial "wow" moment faded out, they were already in the robot acceptance phase.

The Westerner members of an active online community and Japanese owners via Facebook were reached to fill a questionnaire. The questions in this survey asked about the perception of their social robots and the desired improvements in the software. The internal consistency of the answers were verified with Cronbach's α coefficients for four subscales and the exploratory factor analysis mapped the questions into the predefined subscales correctly along with other findings. In overall, the participants rated their robots quite positively after so many years although they wished for many improvements in various functions. The answers were analyzed by four independent variables (gender, age, length of ownership, and culture) and user contribution. Each independent variable had a null hypothesis and the Likert-type ratings were verified by Mann–Whitney or Kruskal–Wallis tests depending on their categories. In gender case, the gender hypothesis (H1) was accepted, the male heavy users tended to have a technological perception of their robots while the female seen a companion in them. The hypothesis (H2) for age groups was rejected, the younger generations were not so tech-savvy and the pensioners did not recognize eagerly these social robots as a companion. The third hypothesis (H3) for culture variable was accepted, and the result was similar to (Bartneck et al. 2007), but the Japanese people were more negative about their Sony AIBOs what contradicts the common stereotype of the robot-loving Japanese society. The length-of-ownership hypothesis (H4) was unexpectedly rejected in a positive manner because the heavy users with over 5 years of ownership rated their bots significantly better. The user contribution question revealed that most heavy users are prepared to make new content expect the young and female. These results were turned into recommendations for social robot design and the limitations of the questionnaire were discussed.

The future work can include similar analysis with heavy users of other social robots which are developed to be a companion for people. The straightforward choice can be the future heavy users of the new Sony ERS-1000 model what revived the AIBO line back to the market in 2018. An interesting question is how we may alter the questionnaire findings if Sony will shift the emphasis in the new robot software from entertainment to companionship. However, executing our survey with the heavy users of other non-animal (e.g., humanoid) robots can be also valuable because their appearances drive different expectations on the uncanny valley and this condition can alter the results. On the other hand, the authors believe that a comparison with a STEM or service robot would not be beneficial since the emotional attachment is essential towards social robots. If this connection is missing, it is unlikely that the owner will use the robot for many years without treating it other than a soulless machine.

Acknowledgements The authors want to say thanks to all enthusiastic members of the aibo-life.org forums to fill out my questionnaire. Special thanks to Christoph Bartneck who shared his valuable research data to draw better conclusions in this paper.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Bartneck C, Suzuki T, Kanda T, Nomura T (2007) The influence of people's culture and prior experiences with AIBO on their attitudes towards robots. AI Soc 21(1–2):217–230

Bernhauerova M (2013) American vs. Japanese management style: which one yields success. MG 201, Introduction to Functions of Management

Coninx A et al (2016) Towards long-term social child-robot interaction: using multi-activity switching to engage young users. J Hum Robot Interact 5(1):32–67

Ezer N (2008) Is a robot an appliance, teammate, or friend? Age-related differences in expectations of and attitudes towards personal home-based robots. Georgia Institute of Technology, PhD Dissertation

Fernaeus Y, Håkansson M, Jacobsson M, Ljungblad S (2010) How do you play with a robotic toy animal?: A long-term study of pleo. In: Proceedings of 9th international conference on interaction design and children ACM, New York, pp 39–48

François D, Powell S, Dautenhahn K (2009) A long-term study of children with autism playing with a robotic pet: taking inspirations from non-directive play therapy to encourage children's proactivity and initiative-taking. Interact Stud 10(3):324–373

Friedman B, Kahn PH, Hagman J (2003) ''Hardware companions?'': what online AIBO discussion forums reveal about the human-robotic relationship. CHI Lett 5(1):273–280

Fujita M (2004) On activating human communications with pet-type robot AIBO. Proc IEEE 92(11):1804–1813

Gockley R, Bruce A, Forlizzi J, Michalowski M, Mundell A, Rosenthal S, Sellner B, Simmons R, Snipes K, Schultz A, Wang J (2005) Designing robots for long-term social interaction. In: 2005 IEEE/RSJ international conference on intelligent robots and systems (IROS), pp 1338–1343

Graaf MM, Ben Allouch S, Dijk JA (2014) Long-term evaluation of a social robot in real homes. In: 3rd international symposium on new frontiers in human-robot interaction (AISB)

Graaf MM, Ben Allouch S, Dijk JA (2016) Long-term acceptance of social robots in domestic environments: insights from a user's perspective. AAAI

Haring KS, Mougenot C, Fuminori ONO, Watanabe K (2014) Cultural differences in perception and attitude towards robots. Int J Affect Eng 13(3):149–157

Haring KS, Silvera-Tawil D, Takahashi T, Velonaki M, Watanabe K (2015) Perception of a humanoid robot: a cross-cultural comparison. In: Proc. of 24th IEEE international workshop on robot and human interactive communication (ROMAN), pp 821–826

Kaplan F (2004) Who is afraid of the humanoid? Investigating cultural differences in the acceptance of robots. Int J Humanoid Rob 1:465–480

Kertész C, Turunen M (2017) What can we learn from the long-term users of a social robot? In: Proc. of 9th international conference on social robotics (ICSR), 2017

Koay K, Syrdal D, Walters M, Dautenhahn K (2007) Living with robots: investigating the habituation effect in participants' preferences during a longitudinal human-robot interaction study. In: The 16th IEEE international symposium on robot and human interactive communication (RO-MAN), pp 564–569

Leite I, Martinho C, Paiva A (2013) Social robots for long-term interaction: a survey. Int J Soc Robot 5(2):291–308

Nomura T et al (2005) Questionnaire–based research on opinions of visitors for communication robots at an exhibition in japan. In: Proc. of IFIP conference on human-computer interaction, pp 685–698

Nomura T, Suzuki T, Kanda T, Kato K (2006) Altered attitudes of people toward robots: investigation through the Negative Attitudes toward Robots Scale. In: Proc. AAAI-06 workshop on human implications of human-robot interaction, pp 29–35

Nomura T, Sugimoto K, Syrdal DS, Dautenhahn K (2012) Social acceptance of humanoid robots in Japan: a survey for development of the Frankenstein syndrome questionnaire. In: Proc. of 12th IEEE-RAS international conference on humanoid robots, pp 242–247

Norman DA (2004) Emotional design: why we love (or hate) everyday things. Basic Books, New York

Salter T, Dautenhahn K, Bockhorst R (2004) Robots moving out of the laboratory—detecting interaction levels and human contact in noisy school environments. In: Proc. of 13th IEEE international workshop on robot and human interactive communication (ROMAN), pp 563–568

Samuels D, Zucco C (2012) Using Facebook as a subject recruitment tool for survey-experimental research. Working paper, Social Science Research Network

Scopelliti M, Giuliani MV, Fornara F (2005) Robots in a domestic setting: a psychological approach. J Univ Access Inf Soc 4(2):146–155

Sung J, Grinter RE, Christensen HI (2010) Domestic robot ecology. Int J Social Robot 2(4):417–429

Yoldas S (2012) A Research About Buying Behaviours of Online Customers. MSc Thesis, University of Roehampton

Zhan K, Zukerman I, Moshtaghi M, Rees G (2016) Eliciting users' attitudes toward smart devices. In: Proc. of the conference on user modeling adaptation and personalization, pp. 175–184

Paper II

Kertész, C., & Turunen, M. (2017). What Can We Learn from the Long-Term Users of a Social Robot? In Proceedings of the 9th International Conference on Social Robotics, 657–665. doi:10.1007/978-3-319-70022-9_65

What Can We Learn from the Long-Term Users of a Social Robot?

Csaba Kertész(✉) and Markku Turunen

Tampere Unit for Computer-Human Interaction,
University of Tampere, Tampere, Finland
csaba.kertesz@ieee.org, markku.turunen@sis.uta.fi

Abstract. Despite the recent technological advances, long-term experiments with robots have challenges to keep the users interested after the initial excitement disappears. This paper explores the user expectations by analyzing the long-term owners of Sony AIBO who have been using these robots for years (heavy users). 78 participants filled an on-line questionnaire and their answers were inspected to discover the key needs of this user group. Quantitative and textual methods confirmed that the most-wanted skills were the interaction with humans and the autonomous operation. The integration with the AI agents and Internet services was important, but the long-term memory and learning capabilities were not that relevant for the participants as expected. The diverse preferences between robot skills led to the creation of a prioritized recommendation list to complement the design guidelines for social robots in the literature.

Keywords: Questionnaire · User expectations · Heavy users · Sony AIBO

1 Introduction

Nowadays more and more social robots are introduced onto the market and the user expectations must be understood for the researchers to execute successful long-term experiments and for the companies to create sustainable business plans. Graaf et al. [6] developed a theoretical foundation to describe the relationship between the owner and its robot over time. In the *adoption* and *adaptation phases*, the first experiences are gained with the robot at home. When the novelty effect fades away and the user expectations are met, daily routines are developed with the robot (*incorporation phase*). After six months, the owner gets emotionally attached to the robot as a personal object (*identification phase*) then the robot is finally accepted for long-term use and the owner becomes a *heavy user*. The ultimate goal in robotics is to reach this acceptance phase and keep the heavy users engaged with the robot for years. Although Nao and other humanoid robots are well-known from the news, their primary users are the research labs while Sony AIBO was the first successful social robot brand offered for private customers. This study focused on the latter robots because their owners are ordinary people forming the eventual target group of the social robotics to bring robots into the mainstream. The Sony robot dogs were so ahead of their time that a significant heavy user base is still reachable long after the discontinuation, therefore, a questionnaire was

A. Kheddar et al. (Eds.): ICSR 2017, LNAI 10652, pp. 657–665, 2017.
https://doi.org/10.1007/978-3-319-70022-9_65

conducted with this user group to identify their long-term expectations. The literature examined the phases before the robot acceptance, but we do not have a broad understanding of the heavy users beyond the identification phase. This paper is a preliminary step to build this knowledge.

The human-robot interaction (HRI) field studied several robots in the past. A general observation was that the users have decreasing interest for the robots after their novelty effect fades away [5, 10]. To avoid losing attraction, social or other engaging capabilities must be identified to create robots for longer use. The long-term interaction was studied by Leite et al. [8] in a survey of exploratory papers in health care, education, work and home settings. They admitted that the reviewed experiments were carried out with limited number of users and the purpose of the longer duration was to let the participants to get comfortable with the experimental conditions. Their results suggested that the people were happy to interact with the social robots for longer periods, but they proposed further analysis to confirm this hypothesis.

Graaf et al. [7] researched with Nabaztag and Karotz robots to create guidelines for better user acceptance. They emphasized the importance of a clear purpose for the robot and the use context because the owners will abandon a robot without utility value. In this way, a truly social personality with interaction skills can differentiate a robot from other gadgets. The authors of [7] also warned the designers that they must consider the mere-exposure effect when the increased familiarity with the robotics technologies will reshape the robot acceptance inside the society over time. The same Karotz experiment was analyzed further in [14] and they found that the users did not reach the acceptance phase mainly caused by the end of novelty, unsatisfied needs, functional replacement (other device) and disappointment in the robot.

The HRI literature showed that the social robots must be designed carefully to engage the users. This study focuses on a domestic social robot, namely, the owners of Sony AIBO robots were analyzed. This product brand included quadruped autonomous entertainment robots which had a behavior-based architecture to exhibit a life-like impression. These robots can walk around the room, interact with the owner and switch between probabilistic state machines to show rich behaviors and engage the owners. Several papers studied AIBO in the past decade, but heavy users were never evaluated directly although Bartneck et al. [1] studied the cross-cultural differences how people perceive AIBO after an interaction session and the on-line forums [4] were analyzed to investigate the relationship between the robots and their owners.

To the best knowledge of the authors, all previous researches in HRI lasted no more than one year and they usually organized short weekly or monthly sessions for the participant together with the robot [2, 3]. Although some relevant results were gained from these past researches, but the subjects in this experiment lived with these social robots day by day for years. This paper explores the expectations of heavy users from a technical perspective what is different from the previous studies which investigated the user perception [6] and the reasons for abandoning robots before the identification phase [14]. Instead of asking the people how they perceive their robot or why they left them behind, the authors addressed in this study what kind of improvements do the participants expect to remain in acceptance phase?

2 Questionnaire

A questionnaire was conducted to get the opinion of people about their Sony AIBO robots. Eight questions asked basic information of the participants (gender, age, home location, profession) and the robot ownership (length, usage frequency, model preference). The usage frequency question ensured that the participants had constant interactions with their robots thus they were part of the target group (heavy users). The following questions were related to the perception of their robots, how they feel about the existing software and which skills must be improved in AIBO. These questions had 9-point Likert-type items (anchors: 1 - Strongly disagree, 3 - Disagree, 5 - Neutral, 7 - Agree, 9 - Strongly agree) and optional text fields were present where the participants could write additional information. A tendency was in the free-form answers that the participants left mostly technical feedback, therefore, two questions (wishes for skills and connectivity options) and the free-form answers were analyzed to characterize the long-term user expectations in this paper.

78 fillings were collected from the members of an English speaking on-line AIBO forum (http://aibo-life.org) what is similar to 70 in [6] and 17 Japanese participants were reached via Facebook ad campaign, similar to [11]. The questionnaire was completed by 57 male and 19 female, since two participants did not reveal their gender, with a ratio 73%/24%, similar to another on-line AIBO questionnaire with 64%/36% in [1] and a robotics questionnaire had 61%/39% in [13]. Although there was no question about the income and the wealth of the participants, the authors assume that this rate can be explained with the higher interest of the men in gadgets [12]. The technical enthusiasm was also reflected in the professions because most owners were e.g. engineers, software developers, technicians (27% for Tech in Fig. 1b) and other occupations were between 1–15% in Fig. 1b.

Since this study focused on robot consumers with more years of ownership, the participants bought and kept these robots after technology acceptance. 14% of these owners were young adult (under 25 years) and their stories on the on-line forum given an insight of their intentions to buy these robots. Either they have got know AIBO in

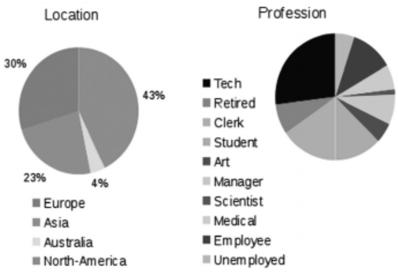


Fig. 1. The home location (a) and profession (b) of the AIBO customers who responded to the questionnaire.

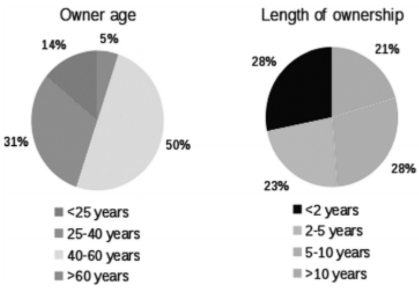


Fig. 2. The age (a) and length of ownership (b) of the participants.

the recent years or they were children during the commercialization of AIBO and they could afford these robots after becoming an adult with income. Figure 2b shows high retention rate for 20% of the participants who kept their robots for more than 10 years, 51% had AIBO between 2–10 years, but 28% possessed AIBO for less than 2 years which is a high rate of newcomers.

3 Result of Quantitative Analysis

Two questions asked the owners about their technical expectations explicitly and both were composed of Likert-type items. Their consistency was verified with Cronbach's α coefficients (0.91, 0.81) thus there was sufficient trust in the overall reliability of the answers.

The average ratings for software improvements in descending order are shown in Fig. 3. Enhancing the dances (5.6), toy-like functions (3.4), the tricks with plastic toys of the robot (ball: 6.7, bone: 6.3) and dog-like behaviors (6.4) had low interest among the participants, most likely, because these features provide the entertainment aspect and the people are more eager to interact with their robots. The exceptionally low ratings of the toy-like function (3.4) can root in the expected anticipation that a robot must be intelligent and it is not a soulless toy. The human-robot interaction capabilities had the highest ratings: speech recognition (7.9), interacting (7.7), distinguish humans (7.7), emotion recognition (7.4), talking (7.1) and playing games (7.0). These skills shape a more valuable emotional connection for the owner towards the robot instead of watching repetitive entertainment behaviors. Although the autonomous features (navigation in rooms: 7.9, object recognition: 7.9) got high ratings, the participants wanted the robot moderately independent (6.8).

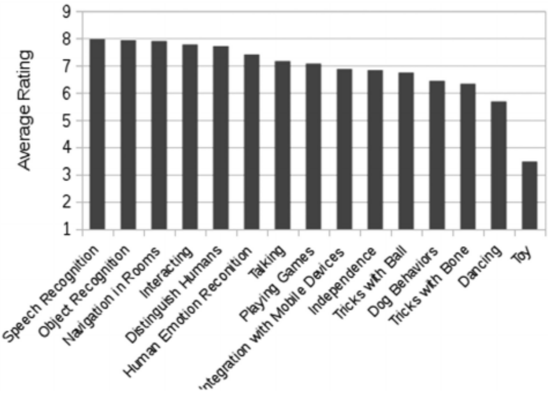


Fig. 3. Average ratings for improvements in the current robot software of Sony AIBO.

The results for connectivity options are shown in Fig. 4. The owners wished most to view robot state (7.3), emotion (7.1) and the camera (7.1), but controlling the robot remotely (5.3) was not interesting. Among the handheld devices, the participants would like to connect their bots to iOS devices (expensive products like robots) with the highest chance (6.6) while Android had a moderate result (5.7), and according to the low market share, Windows devices had the lowest score (4.6). Reading messages (6.2) and emails (6.0) had medium ratings.

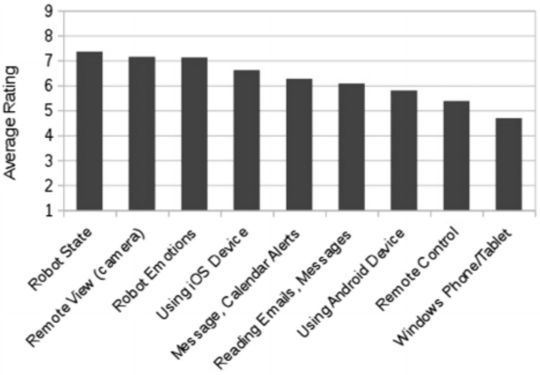


Fig. 4. Average ratings for wished connection options in a new software of Sony AIBO.

4 Result of Free-Form Answers

The participants had the chance to give optional feedback in free form text without any directions. 60% of the Westerners provided this extra feedback with rich statements, nevertheless, the Japanese had a lower profile with 47% and their short answers were infrequent. Despite these data were free text, the participant given consistent answers by emphasizing certain robot skills or pointing out missing capabilities. To analyze the motives, the main points were extracted and counted in all answers like “votes”. A list was created from the results in descending order by votes in Table 1 and the items with less than 5 votes were omitted.

The interaction is essential for social robots. Disappointed Pleo owners reported in [2] that “it would be more important to interact with you then wagging his tail in 28 different ways”. Although AIBO has conversational and interaction skills to some extent, their enhancements were the most requested (1st item in Table 1), similar to the quantitative results in Fig. 3. Suggestions included extensions to the limited vocabulary (50–100 words) and the voice recognition performance was criticized to be far from perfect. With the wave of the AI agents in the smartphones (e.g. Siri, Cortana), people expect to include these technologies in the robots (“It would be great if Aibo could talk through a program like Siri or something.”). The integration with Internet services (5th item) is closely related to the conversational skill, it extends the local AI with live

Table 1. Grouped results of the free form answers by the participants. The main points were extracted from the important statements of the owner feedback. Votes denote the occurrences of a main point in multiple answers.

	Main points	Votes
1	Better conversational and interaction skills	20
2	Autonomous operation	18
3	Richer personality and new content (e.g. motions) over time	16
4	Connectivity options for Apple/Windows/Android devices	16
5	Integration with Internet services (e.g. email, weather, social, Siri)	9
6	More learning capabilities, long-term memory	9
7	AIBO was ahead of its time. Lack of successor	6
8	Good face and object recognition	6
9	More settings to tweak the robot behaviors	6
10	Third-party software is better	5
11	Home automation integration	5
12	Remote control, house surveillance and protection	5
13	Self-charging	5

information from the web ("being able to ask him … what the weather is today, etc. would be great.").

Another prioritized expectation for social robots is to operate autonomously (2nd item), similar to Fig. 3. Albeit people love to interact with their robots, they also like to see their bots wandering around. Genibo, a later Korean robot dog model on the market, was criticized by their consumers that it constantly whines in one place, begs for the attention of its owner and it does not act too much on its own. This feedback exposes the importance of the autonomous activities of the social robots for long-term acceptance.

As the time passes by, people get bored with the same software and they expect new content to keep the amusement with their robots (3rd item). Sony released their AIBO robots with a high price tag and there were no software updates over time. The 3rd party developers got limited chances to build new applications for this brand, nonetheless, the participants praised these software over the official in the 10th item. Nowadays, the success of the mobile app stores and the in-app purchase show that people are prepared to pay for new content if they are worth. The traditional business models can be extended with content purchase or monthly subscriptions to ensure the future commercial success for social robots.

The enhanced connectivity options (4th item) was in tie with the previous item. The participants were reasonable to desire wireless links to their new gadgets after the technological evolution in the past decade. This result is aligned with Fig. 3 where the connectivity skill was ranked after the interaction and autonomous skills.

One surprising finding was that the owners ranked the learning capabilities, memory function, face and object recognition (5th and 8th items) half less important than the top items. Although these abilities are necessary for humans to perceive real intelligence, the utility of a social robot for its owner is focused on building an emotional attachment with the interaction skills.

Some tech-savvy features were ranked to the lowest. The advanced settings for the robot (9th item), the home automation integration (11th item) and the house surveillance (12th item) were present in Table 1, but they had low priority. This outcome suggests for the designers to invent the appearance and tech features according to the specific purpose to maximize the utility.

Some conflicting requests were interesting which are not listed in Table 1. The AIBO software mix the dog-like behaviors with sound effects and the verbal conversations with humans. On one hand, these robots resemble an animal by their appearance and some people wanted to disable the more intelligent features ("I would … turn my ERS7 into a pet dog, no dancing, or talking"). On the other hand, some users would like the opposite, dropping the dog behaviors and including more anthropomorphic features to see a conversational autonomous agent ("Maybe make a new software … with … no dog like actions. And just purely interacts with human speaking.").

The Sony robots had a sophisticated software in the 2000s, but it was far what average people would call artificial intelligence. This chapter given an insight how the heavy users positioned the important skills for their social robot and how the technological evolution influenced these preferences.

5 Discussion

5.1 Social Robot Design

Two papers proposed guidelines to improve the design of social robots for long-term human-robot interactions. Leite et al. [8] presented a good review of this problem with a detailed discussion by accumulating the experiences of different robots in the research literature while [7] expressed their recommendations on a higher level. It was proposed in [7] to create a clear purpose for the robot because this important factor can lead to acceptance by their owners, but we argue that the clear purpose is not enough. If the robot cannot surpass the competing devices in our life in utility value, people will leave and turn to other machines [6]. Based upon the quantitative analysis (Sect. 3) and the ranked textual feedback (Sect. 4), the authors propose the following recommendations in descending priority to complement the past works [7, 8]:

- Design the robot appearance according to its capabilities to avoid the uncanny valley [9]. Sony AIBO was successful because it resembled an animal, walked around and had interactions with humans according to the ads. The owners expected less intelligence from these robot dogs than a humanoid robot [5]. Pleo owners were disappointed in [2] when they realized that their robots have legs, but they cannot walk without a trigger.
- The interaction and conversational skills were found the most important by the users. This is in accordance with the third recommendation in [7], the perceived sociability is a key factor for the users to accept the robot.
- Integrating the latest web services and conversational agents into the robot enhances the user experience. When intelligent applications (e.g. Siri, Alexa) are available in handsets or computers, people expect to have similar built-in skills in their robots.

- The owners expect autonomy, self-charging from a mobile robot as well as a remote application to interact with the robot and check its status.
- The people gets bored quickly with the repetitive behaviors [10], but if the robot is attractive enough, the owners keep the robots for a long time. The participants of this experiment verify this assumption. However, the users expect new behaviors constantly (Sect. 4) and this opportunity can be turned into a business strategy for social robots to charge regular fee per content update, similar to the mobile applications. This approach with a broad userbase can create a sustainable revenue for a robotics company.
- The learning and memory skills are hard problems in the artificial intelligence. Leite et al. [8] stated that the benefits of memory is unclear in the long-term interaction, nevertheless, AIBO users explicitly asked this skill. Although this feature is important, it can be prioritized less than the communication and interactions skills according to Sect. 4.
- The target group of the social robots is broad from the teenagers to the pensioners (Fig. 2) although 50% were between 40–60 years.

5.2 Limitations

Despite of the participants were recruited on a special internet forum and Facebook, only 78 active Sony AIBO owners were reached, but the authors believe that sample size was reasonable compared to 230 in [1] and 41 in [13] considering that conducting our survey was long after the product discontinuation. The sampling was not representative for the general public, but the participants could provide a good indication about the typical users of entertainment robots and even beyond this group since Bartneck et al. found in [1] that owning a Sony AIBO did not result significantly different scores on their NARS questionnaire.

Since these robots were commercial, this study was essential to analyze the heavy users of an expensive robot from the market. The robots in past experiments were given to participants on a voluntary basis for free [2, 7].

5.3 Conclusion

The Westerner and Japanese heavy users of Sony robot dogs were studied with a questionnaire in this paper after 10 years of the product discontinuation. Since these people owned their robots for years after the initial "wow" moment faded out, they were already in the robot acceptance phase. Despite Sony AIBO was ahead of its time, it exhibited several mechanical and software limitations, and definitely, the heavy users were not satisfied with the robot capabilities after many years without software updates. Both the quantitative analysis and the free-form text answers suggested that the most-wanted improvement was the interaction skills with humans, followed by the autonomous operations. The participants was not interested too much in the entertainment aspects, remote control or self-charging, but they would like to connect their robot with handheld devices and modern Internet services. It was surprising that the learning capabilities and long-term memory were moderately important for the users.

After the questionnaire analysis, recommendations were written for social robot design to complement the initial guidelines in the literature [7, 8].

The future work can include similar analysis with heavy users of other robots and it can be worth to compare our results with different robot appearances or personalities.

Acknowledgements. The authors want to say thanks to all enthusiastic members of the aibo-life.org forums to fill out our questionnaire.

References

1. Bartneck, C., Suzuki, T., Kanda, T., Nomura, T.: The influence of people's culture and prior experiences with AIBO on their attitudes towards robots. J. AI Soc. **21**(1–2), 217–230 (2007)
2. Fernaeus, Y., Håkansson, M., Jacobsson, M., Ljungblad, S.: How do you play with a robotic toy animal?: a long-term study of Pleo. In: Proceedings of 9th International Conference on Interaction Design and Children, pp. 39–48 (2010)
3. François, D., Powell, S., Dautenhahn, K.: A long-term study of children with autism playing with a robotic pet: taking inspirations from non-directive play therapy to encourage children's proactivity and initiative-taking. J. Interact. Stud. **10**(3), 324–373 (2009)
4. Friedman, B., Kahn, P.H., Hagman, J.: "Hardware companions?": what online AIBO discussion forums reveal about the human-robotic relationship. CHI Lett. **5**(1), 273–280 (2003)
5. Gockley, R., Bruce, A., Forlizzi, J., Michalowski, M., Mundell, A., Rosenthal, S., Sellner, B., Simmons, R., Snipes, K., Schultz, A., Wang, J.: Designing robots for long-term social interaction. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1338–1343 (2005)
6. Graaf., M.M., Ben Allouch, S., Dijk, J.A.: Long-term evaluation of a social robot in real homes. In: Proceedings of 3rd International Symposium on New Frontiers in Human-Robot Interaction (AISB) (2014)
7. Graaf, M.M., Ben Allouch, S., Dijk, J.A.: Long-term acceptance of social robots in domestic environments. In: AAAI (2016)
8. Leite, I., Martinho, C., Paiva, A.: Social robots for long-term interaction: a survey. Int. J. Soc. Robot. **5**(2), 291–308 (2013)
9. Mori, M.: The uncanny valley. Energy **7**(4), 33–35 (1970)
10. Salter, T., Dautenhahn, K., Bockhorst, R.: Robots moving out of the laboratory—detecting interaction levels and human contact in noisy school environments. In: Proceedings of 13th IEEE International Workshop on Robot and Human Interactive Communication (ROMAN), pp. 563–568 (2004)
11. Samuels, D., Zucco, C.: Using Facebook as a subject recruitment tool for survey-experimental research. Soc. Sci. Res. Netw., 1–20 (2012)
12. Yoldas, S.: A research about buying behaviours of online customers. MSc thesis, University of Roehampton (2012)
13. Haring, K.S., Mougenot, C., Fuminori, O.N.O., Watanabe, K.: Cultural differences in perception and attitude towards robots. Int. J. Affect. Eng. **13**(3), 149–157 (2014)
14. Graaf, M.M., Allouch, S.B., Dijk, J.A.: Why do they refuse to use my robot?: reasons for non-use derived from a long-term home study. In: Proceedings of International Conference on Human-Robot Interaction, pp. 224–233 (2017)

Paper III

Kertész, C., & Turunen, M. (2018). Common Sounds in Bedrooms (CSIBE) Corpora for Sound Event Recognition of Domestic Robots. Intelligent Service Robotics, 11(4)335–346. doi:10.1007/s11370-018-0258-9

Intelligent Service Robotics (2018) 11:335–346
https://doi.org/10.1007/s11370-018-0258-9

ORIGINAL RESEARCH PAPER

Common sounds in bedrooms (CSIBE) corpora for sound event recognition of domestic robots

Csaba Kertész[1] · Markku Turunen[1]

Received: 2 September 2017 / Accepted: 22 August 2018 / Published online: 28 August 2018

Abstract
Although sound event recognition attracted much attention in the scientific community, applications in the robotics domain have not been in the focus. A new database was published in this paper and classifiers were evaluated with this dataset to guide the future practical developments of domestic robots. A corpus (CSIBE-RAW) was collected from the internet to build acoustic models to recognize 13 sound events and omit ambient sounds. As a case study, CSIBE-RAW was rerecorded in four room settings (CSIBE-AIBO) to create reverberation-tolerant classifiers for a Sony ERS-7. After eight classifiers were reviewed, the convolutional neural network achieved the best accuracy (95.07%) after multi-conditional learning and it was suitable for real-time classification on the robot. The effects of lossy audio codecs were studied, lossy encoder-tolerant audio statistics were specified for the feature vector and the Ogg Vorbis encoder with 128 kbit VBR was found superior to store big data and avoid any significant accuracy loss with the compression ratio 1:8.

Keywords Domestic robots · Recognition · Indoor audio corpus · Sony AIBO · Deep learning

1 Introduction

The contextual interpretation of the environment involves the fusion of multiple cues for the human beings [15]. The speech recognition [43] and music annotation [24] have been traditional research fields in robotics to reach human-level performance in auditory perception. Sound event recognition is a relatively new research field in the last decade that shifted the interest from the anthropomorphic bias to a more natural point of view of the scene. This paper introduces a new corpora for isolated sound event recognition for indoor robotic applications and some common problems are highlighted with possible solutions to build robust acoustic models.

The only sound event database for robotics was published in Maxime et al. [20]. The NAR Dataset was recorded with a Nao humanoid robot from Aldebaran Robotics in a kitchen and contained 22 sound events as well as 20 English words. The average signal-to-noise ratio (SNR) of the recordings was 15 dB because of the noisy fans inside the robot body and the SVM classifier achieved 91.5% accuracy after tenfold cross-validation despite the challenging conditions. This result was reached with file-averaged feature vectors and the model was not evaluated with unseen data (unused data during training and cross-validation).

The Acoustic Event Dataset (AED) [28] was recorded in a smart room environment and Gaussian mixture model (GMM) was trained with a 600 ms sliding data window. The classifier was tested with unseen data and it distinguished 11 events with 87% accuracy from their small database. The unknown events were not modeled in their system, but the silence was a separate class.

The IEEE Detection and Classification of Acoustic Scenes and Events (IEEE-DCASE) Challenge was organized in 2013 to establish an international competition for identifying sound scenes and events. The DCASE-OL dataset with 16 events was dedicated for event detection in real office environment [35]. The best system in the challenge had 61% frame-based precision on unseen data without modeling the background sounds.

Beltrán et al. [2] proposed a novel sound event recognition method with temporal histograms of Mel-based Multi-Band Spectral Entropy Signature coefficients and they reported better results than MFCC-based SVM classification with source separation (non-negative matrix factorization). The

✉ Csaba Kertész
csaba.kertesz@ieee.org

Markku Turunen
markku.turunen@uta.fi

[1] University of Tampere, Kalevantie 4, 33100 Tampere, Finland

ambient sounds were not modeled, but their approach can detect the mixture of two events without any source separation technique. Their CICESE corpus contained several reusable datasets, but most of them were incomplete compared to the dataset description in their published paper which makes any comparison hard with their results.

Unlike the previous works with event classes, the Computational Hearing in Multisource Environments in Home (CHiME-Home) database [14] was annotated on a higher granularity for speech, human activity, television and household appliances. 4-s-long audio chunks were allowed to hold multiple labels and a GMM classifier obtained 89% accuracy after tenfold cross-validation. Some event classes represented ambient sounds (television, household appliances) what is an advantage, but on the other side, the GMM classifier was not evaluated with unseen data.

To mention an outdoor example, a work [33] invented a new taxonomy for urban sound classification and their dataset (UrbanSound8K) which had 18.5 h of audio with 10 events. Temporal statistics complemented the feature vectors and they found the 4-s-long sliding window optimal. The model performance was estimated to be 69% with tenfold cross-validation of random forest (RF) and SVM classifiers.

Usually, the uninteresting sound events and the background noises were not modeled in the past works [2, 33, 35] unlike in [14]. The new corpora (CSIBE) in this paper are specialized to the indoor robotics applications and an event class is dedicated to represent the auditory background with appliance sounds, object and human related noises. Free audio files were downloaded from the internet into CSIBE-RAW in order to have a clear licensing situation for the further usage. Creating universal acoustic models is a challenge because of the different noise levels and microphone characteristics. To provide a practical example for robots, the collected sounds were recorded again by replaying through a high-quality speaker and capturing with the stereo microphones of a Sony ERS-7 robot dog into CSIBE-AIBO dataset. A baseline sound event recognition system was developed and CSIBE datasets were evaluated with tenfold cross-validation and unseen data. CSIBE-AIBO had to be stored in lossy audio format, therefore, the effect of Ogg Vorbis and MP3 codecs were examined before drawing the conclusion at the end of the paper.

2 CSIBE corpora

The available sound event corpora contain events of specific scenarios like smart room [28], office [35], kitchen [20], urban area [33] and they have been built without modeling the ambient sounds. This paper proposes that modeling the ambient sounds and audio sources in multiple locations are important to shorten the gap between research and practical applications. The authors could not find a free or commercial audio corpus for indoor applications which satisfied two criteria. Firstly, the desired dataset has a separate event class for *ambient sounds* that should not take the attention of the robot, secondly, the samples provide high intraclass variability. Therefore, a corpus was assembled from free online databases, public research datasets, our own recordings and it was shared with the scientific community (https://doi.org/10.5281/zenodo.1243714). The license of each sound sample was included in the package to enable the reuse with a clear legal status of the research data. CSIBE consists of typical sound events in bedrooms/living rooms as people interact with social robots in these places at home. The database have two parts:

- *CSIBE-RAW* Human speech and other events were collected from the internet in this dataset and complemented with new recordings. The samples had excellent and clear sound quality, they were stored in mono WAV format with 16 bit depth, 44.1 kHz sampling rate. All files were labeled according to the sound event type.
- *CSIBE-AIBO* The samples of CSIBE-RAW were played back through a speaker and recorded with the robot microphones. One mono microphone is located in each ear of the robot. The details about the recording conditions are described in Sect. 2.2.

2.1 CSIBE-RAW

The core part of CSIBE was collected from public internet sources (freesound.org, AED [28], DCASE-OL-IEEE DCASE Challenge [35], NAR Dataset [20]) and new recordings were done. The sample file counts of 14 sound events in the database were balanced around 300 except the ambient sound and speech (Fig. 1). The first was overrepresented to have a strong class for the uninteresting events, the latter was important for reliable speech detection in human–robot interactions. Further characteristics of CSIBE-RAW can be noted in Fig. 1 when the sample counts (blue lines) are compared to the sample size (red line). When those two lines are close to each other for a certain event (e.g., piano, bell, vacuum cleaner, traffic), the average duration of the samples are close to 1 s, but short events (100–200 ms) cause a shorter red line (e.g., flatulence, parrot, ambient sound) because the same amount of data is divided among more samples.

The overall size of CSIBE-RAW is comparable to the indoor sound event databases in the literature:

- AED for smart rooms has 11 events (350 MB), 213 samples.
- NAR [20] for kitchen scene with Nao robot has 42 events, 831 samples (42 MB).

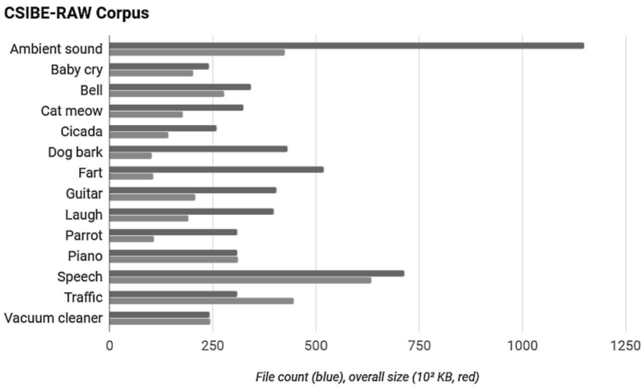


Fig. 1 Sample counts for each sound event in CSIBE-RAW

– DCASE-OL [35] for office scene has 16 events (398 MB in 16 bit WAV), 1280 samples.

Although more events are in NAR (42) and DCASE-OL (16), CSIBE-RAW has a much larger sample set (5954 samples). Each event in CSIBE-RAW was recorded with multiple microphones and sound sources to ensure the high intraclass variability and the higher chance for interclass correlation. For the diverse sources, the speech event is an good example in which male, female and children voices in various languages (English, Spanish, Hungarian, French and Japanese) are represented. Enough samples were available for all events on the internet except flatulence and parrot. The church bell and traffic noises are environmental sounds, but they were included in the database because they are audible inside, similar to cicada, that is a common insect in Asia with high pitch sound. The ambient sounds include insignificant events for robots and short events which are challenging for even humans without visual cue: door openings, key falls, knocks, moving chairs, keyboard clicks, breathing, throats, coughs and steps.

CSIBE-RAW was randomly split into a training ($\mathrm{CR_T}$) and a validation set ($\mathrm{CR_V}$). Unlike the usual process to create a larger training set, the authors put 26% of the samples to the training and 74% to the validation set in order to prove the good discriminative power of the classifiers. The split is denoted in the following brackets ($\mathrm{CR_T}/\mathrm{CR_V}$) for each event: *ambient sound* (326/824 samples), *baby cry* (39/201 samples), *church bell* (70/273 samples), *cat meow* (62/263 samples), *cicada* (39/221 samples), *dog bark* (149/282 samples), *flatulence* (226/292 samples), *guitar* (109/296 samples), *laugh* (154/244 samples), *parrot* (111/198 samples), *piano* (57/253 samples), *speech* (125/588 samples), *traffic* (32/277 samples) and *vacuum cleaner* (32/211 samples), in overall 5954 samples. The audio data durations of the classes were balanced in $\mathrm{CR_T}$ thus there were many short flatulences in $\mathrm{CR_T}$, but fewer samples of cat meows, traffic and vacuum cleaner.

One file in the database corresponds to one sound event sample where the silent audio chunks were removed from. In this way, the database is easier to process because there are no additional files for labeling.

2.2 CSIBE-AIBO

The standard datasets are often not suitable to develop practical acoustic models for robotics because the training data must incorporate microphone dynamics, various noise levels and reverberant conditions. CSIBE-AIBO attempts to step forward in this direction to show how a base model can be developed by rerecording CSIBE-RAW in multiple settings.

Sony ERS-7 has two “ears” with microphones and these devices feature 16 kHz sampling rate in 16 bit depth. This robot is an embedded platform, therefore, the recording quality is not so clear like a Zoom H1 or H2 recorder. CSIBE-RAW samples were played back with a high-quality speaker (Audio Pro Addon One) in silent rooms from different locations relative to the robot:

- Reverberant room, speaker was 1 m away, 30° counterclockwise to the head.
- Non-reverberant room, speaker was 1 m away, 30° counterclockwise to the head.
- Reverberant room, speaker was 3 m away, 1 m high, 180° clockwise to the head.
- Non-reverberant room, speaker was 3 m away, 1 m high, 180° clockwise to the head.

CSIBE-RAW samples were recorded in these room settings with stereo microphones which resulted eight times more data compared to the mono CSIBE-RAW samples. The new sounds were affected by the low-end input quality, reverberation, the recording distance, the microphone displacement relative to the source direction and the microphone self-noise. CSIBE-AIBO was separated to training and validation sets using the same partitions as CSIBE-RAW. $\mathrm{CA}'_{\mathrm{T}}$ contains the rerecorded samples of CR_{T} in the first setting, CA_{T} in all settings. $\mathrm{CA}'_{\mathrm{V}}$ and CA_{V} were generated from CR_{V} with the same declaration, respectively.

Transferring the audio data from the robot was a challenge because CSIBE-RAW has more than an hour of samples and this procedure had to be automated. The internal storage of the robot could not be used to store the rerecorded samples by reasons of being small (max. 128 MB) and slow read/write speeds. The other option was to transfer the recordings from AIBO via the built-in wireless card with low throughput (appr. 30 KB/s). The authors found the best compromise with encoding the recorded audio in lossy format and sending the compressed data to a PC on the same WLAN. The SNR of the sound events in CSIBE-AIBO varied between 8.31 and 16.15 dB (average: 10.85 dB) which was a bit lower than the noisy NAR Dataset (15 dB).

The target was to find a lossy codec which did not affect the classifier accuracy if either the training or the validation set is transcoded. *Full transcoding* denotes when both the training and the validation set are transcoded with the same lossy codec before the model building and evaluation processes. This is relevant for storing large audio databases in the fraction of the original disk space and using these big data to train and test deep neural networks without performance degradation. *Half transcoding* means transcoded training set and unaltered validation set. This evaluation step ensures that DNN models built with full transcoding can be deployed on consumer devices where the model will receive uncompressed audio from a real microphone.

Two popular lossy compression algorithms were examined: Ogg Vorbis and MP3. The Ogg Vorbis transcoding was implemented with libvorbis,[1] the MP3 encoding with libmp3lame[2] and the MP3 decoding with LAME's mpglib version. The effects of lossy encoding has been studied in the literature for speech [4, 29, 32] and music [40, 41] classifications. These papers examined only the MP3 encoding while both Ogg and MP3 codecs are reviewed here for sound event recognition. Some past works built the acoustic model with uncompressed audio and the validation set was transcoded with lossy codecs to check for any loss in performance [4, 5], but half transcoding was applied in [23] and full transcoding on both the training and validation sets in [25, 40, 41]. Minimum 32 kbit MP3 profile was sufficient to avoid performance decrease for speech recognition in [4, 23] and 64 kbit in [32]. Uemura et al. [40] found 32 kbit VBR enough for chord recognition while Urbano et al. preferred at least 160 kbit CBR MP3 encoding [41]. The encoding parameters were explored during the initial experimentation and they are discussed in Sects. 4.5 and 4.6.

3 Recognition system

Features must be extracted from the audio to train a classifier. The most popular features in the literature are the Mel-frequency cepstrum coefficients (MFCC) [2, 22, 33]. In this paper, the audio data were framed by a sliding Hann-filtered window (32 ms) with 33% of overlap, fast Fourier analysis was performed to extract harmonic spectrum, spectral peaks and 26 MFCCs for each frame. The first MFCC coefficient was dropped because it measures the signal loudness and the remaining were added to the feature vector. There is no clear consensus in the literature about the ideal MFCC count, some earlier studies employed 13 MFCC components [2, 11, 37], but other works included 15 (Phan et al. [26]), 16 (Mesaros et al. [22]), 20 (Ruiz-Martinez et al. [31]), 26 (Salamon et al. [33]) and 40 (Nouza et al. [25]).

The feature extraction was done with the libxtract library [7] in C++and the implementation details of each audio statistic can be found in the github repository.[3] The following 23 statistics were calculated to complement 25 MFCC to 48 features:

- Audio data frames: standard deviation, maximum, min—max range, kurtosis, fundamental frequency, non-zero count, average deviation, variance and zero crossing rate.
- FFT spectrum: pitch of Harmonic Product Spectrum analysis, irregularity [18], centroid, variance and standard deviation.
- Bark coefficients: loudness.
- Peak spectrum: standard deviation, partials count (non-zero component count) and centroid.
- Harmonic spectrum: arithmetic mean and tristimulus [30].
- MFCC frames: minimum, arithmetic mean and standard deviation.

These statistics were selected by sequential forward floating feature selection and an iterative examination of the feature importances with cross-validation in the initial experiments. Two features (spectral crest, variance of spectral harmonics) were removed since they were sensitive to lossy encoding. Eventually, the feature vector had small compu-

[1] https://xiph.org/vorbis/.

[2] http://lame.sourceforge.net.

[3] https://github.com/jamiebullock/LibXtract.

tational cost (1 ms) on the robot and it was robust to lossy audio codecs.

The temporal frame integration (superframes, bag-of-words) was tried without satisfactory results in the initial experiments, therefore, majority voting was used for temporal smoothing to do the classification of each sample file in the next sections. When a label must be associated with a file, feature vectors are extracted, predicted with a classifier and the label with the most predictions is voted to be the final.

4 Experimental results

4.1 Classifier comparison

Most classifiers were implemented with the Machine Learning module of OpenCV [6]. Other libraries provided implementations for maximum entropy (ME) [1, 39], convolutional neural network (tiny-dnn[4]) and SVM with linear kernel [19]. The latter was chosen because of the SVM codes in OpenCV uses an old fork of libsvm with custom modifications and several users reported reduced performance with linear kernel behind libsvm. The linear SVM with Dual Coordinate Descent Method [17] in Dlib [19] provided better accuracy than the SVM in OpenCV for the authors with this dataset. The SVMs used $C = 0.1$ parameter while ME was regularized with an Orthant-Wise Limited-memory Quasi-Newton Optimizer ($\mathrm{L1} = 0.00001$). The hyperparameters for the decision tree and random forest were $\mathrm{Tree}_{\mathrm{depth}} = 20$, $\mathrm{Forest}_{\mathrm{size}} = 20$ and the minimal sample count for node split was set to 100.

Several earlier studies focused on deep neural networks for sound event recognition [8, 10, 16]. Some explored the input features [13, 16, 21] for the networks, some compared different network topologies [21, 27]. This paper employs a simple convolutional neural network (Fig. 2) without automatic feature extraction which can be deployed to embedded systems. The first two layers in the proposed neural network were fully connected with 200 units, one convolutional layer had 9×1 kernel with stride 1 and the last fully connected layer contained 100 neurons. The fully connected layers had leaky rectified linear activation function and the convolutional layer had tanh. The CNN training was executed for 50 epochs with adaptive gradient method (adagrad), mean squared error loss function and batch size 64. Although the state-of-the-art works in the audio literature use recurrent neural networks (RNN) for classification problems, the authors favored the CNN architecture. This decision was made because CNNs are more accessible in the research software and RNNs provide only marginal improvements over CNNs [42].

Each hyperparameter was specified with parameter search on preliminary data during the initial experimentation. The features were rescaled to [0, 1] for CNN. Other classifiers got the data after standardization before training and prediction.

The previous datasets were evaluated with cross-validation (CV) [2, 35, 37] that estimates the model accuracy, but the proper validation is done with unseen data. The tenfold cross-validation was done with the training sets of CSIBE (CR_T, CA'_T) in this paper. Then the model generalization was explored by building models with these sets and evaluating them with CR_V or CA'_V.

Figure 3 shows the CV of eight classifiers with the training set of CSIBE-RAW (CRT) and the rerecorded version (CA'_T) from CSIBE-AIBO where all results were calculated with majority voting. Apart from the standard, frame-based evaluation, aggregated frames [3] were computed by replacing the original frames with temporally calculated mean and standard deviation of every 9 frames with 30% overlap ($\mathrm{CR}_{\mathrm{T,a9}}$, $\mathrm{CA}'_{\mathrm{T,a9}}$). This technique results smaller training set size with negligible computational costs and the accuracy was improved expected compared to the frame-based evaluation in [37]. As we can see, the aggregated frames enhanced the performance (even columns are higher) in Fig. 3 except $\mathrm{SVM}_{\mathrm{RBF}}$, KNN and DT what was a bit surprising and the authors do not know the reason why these classifiers would be sensitive for this method. The aggregation parameters were selected empirically, but varying the frame count or the overlap size did not have substantial changes on the effects during the initial experimentation. CNN was the top performing classifier in all cases, nevertheless, KNN, DT, RF provided solid accuracies over 80%. Because of $\mathrm{SVM}_{\mathrm{RBF}}$ did not reach the performance of other classifiers, it was left out from the further analysis.

SVM can collapse if the training dataset is too big, the training time can increase rapidly and performance is degraded. SVM can perform better if the frame count is reduced with sliding window transformations [20, 37]. $\mathrm{SVM}_{\mathrm{RBF}}$, NB on Fig. 3 had the biggest improvements among the other classifiers when the cross-validation was done with aggregated frames ($\mathrm{CR}_{\mathrm{T,a9}}$, $\mathrm{CA}'_{\mathrm{T,a9}}$) which reduced the training set sizes to 1/3. Although the aggregation yielded the least improvement for CNN, but its training time was reduced by 1/3 what is important to speeding up the slow training of deep learning networks.

The past studies used various methods to check their datasets and a natural point is how the baseline system (BS) in this paper relates to the literature. Two public corpora were tested by BS with the same preprocessing steps of the original studies, but distinct classifier implementations and audio statistics. The dataset A of CICESE was cross-validated with a hidden Markov model (HMM) to 98% (F-score) in [2] while BS reached 99.2% accuracy with KNN, 98.8% with DT and 98.3% with RF. The NAR dataset was cross-validated

[4] https://github.com/tiny-dnn/tiny-dnn.

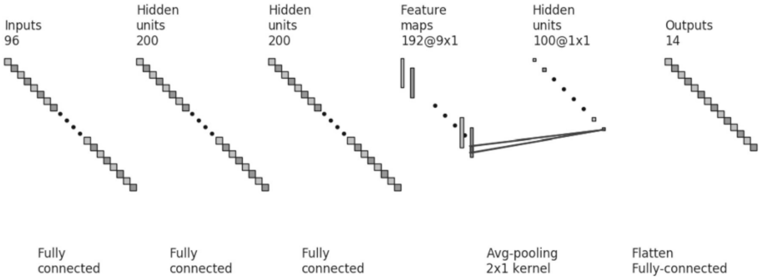


Fig. 2 The used convolutional neural network topology in this paper This figure was generated by adapting the code from https://github.com/gwding/draw_convnet

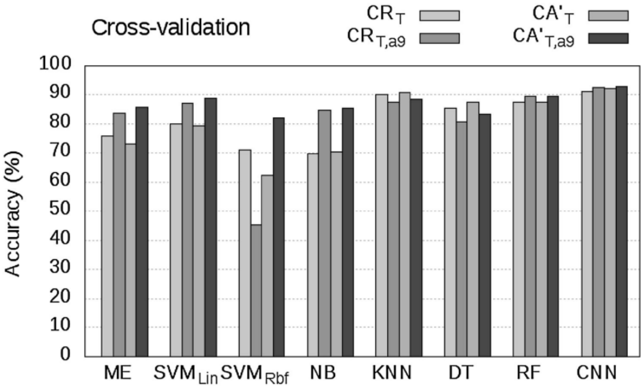


Fig. 3 Cross-validation results for several classifiers over the training set (CR_T) of CSIBE-RAW (red) and CA'_T of CSIBE-AIBO (blue). The lighter colors (odd columns) were the frame-based votes and the aggregated frames were used in the darker (even columns), but the final results were calculated with majority voting. Classifiers: maximum entropy (ME), support vector machine with linear kernel (SVM_{Lin}), support vector machine with radial basis function (SVM_{Rbf}), naïve Bayes (NB), k-nearest neighbors (KNN), decision tree (DT), random forest (RF) and convolutional neural network (CNN) (color figure online)

with KNN to 88.4% and SVM to 91.5% in [20] while BS had 91.41% with KNN and 92.1% with SVM_{Lin}. The above results showed that the recognition system of this paper provides the same or slightly better performance compared to [2] and [20].

4.2 Model evaluation

The cross-validation estimates the model accuracy with unseen data, but it can lead to misunderstandings about the generalization power. Figure 4 shows the model evaluations for seven classifiers which were selected after the cross-validation in Sect. 4.1. Models were built with the aggregated frames of the previous training sets ($CR_{T,a9}$, $CA'_{T,a9}$) and they were evaluated with the aggregated frames of their validation sets ($CR_{V,a9}$, $CA'_{V,a9}$) to identify any difference in the performance with unseen samples. The aggregated frame-based accuracies (lighter columns: CR_{af}, CA'_{af}) varied between 70 and 90%, the majority voting enhanced these results up to 90–96% (darker columns: CR_{mv}, CA'_{mv}). All classifiers were satisfactory after majority voting, but CNN was again the best in every situation. The actual model accuracies were underestimated by the cross-validation because all CR_{mv} and CA'_{mv} evaluations in Fig. 4 were over 90% while almost none of the cross-validations of the same classifiers reached 90% in Fig. 3.

CSIBE-RAW contains the collected samples from the internet and CSIBE-AIBO the rerecorded versions. The original sounds achieved higher frame-based results in all cases (CR_{af} vs. CA'_{af} in Fig. 4) because the rerecorded sounds in CA' had worse SNR ratio, but the majority voting smoothed this disadvantage and every CA'_{mv} was on the same level with CR_{mv}. The classifiers delivered accuracies over 90% with majority voting regardless of the datasets (CR vs. CA'). The next subsection will examine the learning challenges in real-world applications.

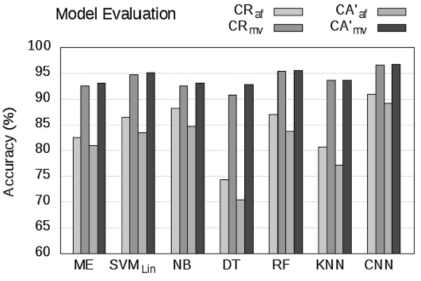


Fig. 4 Model evaluation of seven classifiers over the validation set (CR_V) of CSIBE-RAW (red) and CA'_V of CSIBE-AIBO (blue). The lighter colors (odd columns) were the aggregated frame-based evaluations and majority voting was used in the darker (even columns) in addition (color figure online)

4.3 Multi-conditional learning (MCL)

The room reverberation and the microphone dynamics alter the feature vector and may raise the recognition error significantly. In MCL, the classifier is trained with samples with different distortions and the built model will be robust to these conditions. Training and validation sets were constructed in Sect. 2.2 for MCL, CA'_T contained rerecorded samples of CR_T from one room setting. However, the same sample files rerecorded in all four settings were included in the multi-conditional training set CA_T with multiple reverberant and SNR conditions.

In [22], the isolated sound events were recognized with a GMM-HMM model and the system had 53% accuracy for clean samples, 47% for 10 dB SNR, 38% for 5 dB SNR and it decreased to 28% for 0 dB SNR. Ruiz-Martinez et al. implemented SVM for environmental sounds [31] and their model achieved 89% accuracy for clean samples, 85% for 10 dB SNR, 79% for 5 dB and 71% for 0 dB. These earlier works had considerable loss in the performance by adding artificial noise which can be handled with some solutions. Artificial noise was not used in this work, the samples in CA_T comprised real SNR values between 8.31 and 16.15 dB to imitate the everyday conditions. The model adaption, signal enhancement and feature compensation can make the system more robust against noise [13]. Multi-conditional learning is used in this paper to build models that are tolerant for lower SNR, reverberant and non-reverberant conditions. On one hand, this learning method requires large training set, on the other hand, the authors wanted to avoid synthetic training data that is sometimes not applicable outside the laboratory environment. The CSIBE-AIBO corpus was built according to these principles (Sect. 2.2) although the rerecording in four settings was a time consuming process.

Table 1 Model evaluation without multi-conditional learning

Training set	Validation set	DT	RF	SVM_{Lin}	CNN
$CR_{T,a9}$	$CR_{V,a9}$	90.78	95.41	94.69	**96.45**
$CA'_{T,a9}$	$CA'_{V,a9}$	92.83	95.57	95.07	**97.54**
$CR_{T,a9}$	$CA'_{V,a9}$	65.36	74.84	79.20	**79.88**
$CR_{T,a9}$	$CA_{V,a9}$	38.36	44.28	**53.22**	52.02
$CA'_{T,a9}$	$CA_{V,a9}$	62.30	62.21	64.62	**66.49**

Decision tree, random forest, linear SVM and convolutional neural network performances with different training and validation sets. All results were calculated with aggregated frames and majority voting. The first two rows contain accuracies from Fig. 3

The best accuracy in each row is bold

The aforementioned problems are analyzed in Table 1 how DT, RF, SVM_{Lin} and CNN classifiers performed in model evaluation when the training and the validation sets were varied. The first two rows represent the baseline and come from Fig. 4. The model of the first row was built and evaluated on the clean samples of CSIBE-RAW, all accuracies were over 90%. Similar results were achieved with rerecorded sets in the second row because the recording environment altered both the training and the validation sets.

When the clean training set ($CR_{T,a9}$) was tested against the rerecorded validation set in one room setting ($CA'_{V,a9}$), the accuracies were decreased by 15–25% in the third row compared to the first. Evaluating $CR_{T,a9}$ with the rerecorded validation set in all four settings (fourth row), the performance was dropped even more (41–52%) compared to the first row. These results confirmed that models built on the original samples in CSIBE-RAW could not generalize for the recording conditions of CSIBE-AIBO. The last row in Table 1 contains the results for rerecorded training set in one setting ($CA'_{T,a9}$) and rerecorded validation set in four room settings ($CA_{V,a9}$). Despite the both sets were affected by the robot microphone and the reverberation, the accuracies were as low as 62–64% since $CA_{V,a9}$ was altered by all four settings.

To summarize the findings:

- CNN delivers the best performance almost every time without MCL (except the fourth row in Table 1).
- The more challenges the validation set have (reverberation, lower SNR), the more the accuracies decrease.
- Deep neural networks cannot handle these problems with hand-crafted features. (Moving the feature extraction to autoencoders can be a solution if large amount of training data and GPU power are available.)

As a consequence, when big data is not accessible, multi-conditional learning can be a solution. McLoughlin et al. [21] trained deep neural networks with spectrogram image features from different noise conditions. Their models deliv-

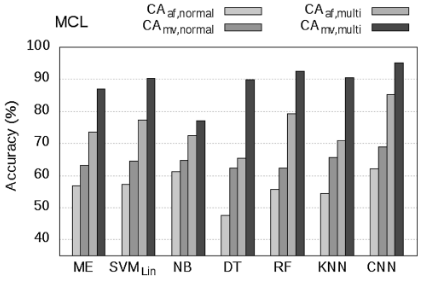


Fig. 5 Model evaluation with and without multi-conditional training. The validation set is $\mathrm{CA}_{\mathrm{V,a9}}$ of CSIBE-AIBO in all cases, but the training set is $\mathrm{CA'}_{\mathrm{T,a9}}$ for the red columns and $\mathrm{CA}_{\mathrm{T,a9}}$ for blue columns. The lighter colors show the aggregated frame-based evaluation and the darker represent the results after majority voting (color figure online)

ered similar accuracies for the clean and 20 dB SNR testing samples though the accuracy decreased by 1–6% in the case of 10 dB SNR. In Terence et al. [37], when the feature vectors based on MFCC components were trained to a GMM, their model had 67.40% accuracy without MCL and 95.12% with MCL. Dennis [13] had two systems based on SVM and HMM and dropped 20–90% accuracy without MCL, but the degradation was reduced to 2–30% under 0–20 dB SNR conditions with MCL. According to these earlier works, multi-conditional learning is an effective method to deal with the negative effects of different noise levels.

$\mathrm{CA}_{\mathrm{af,normal}}$ and $\mathrm{CA}_{\mathrm{af,multi}}$ (red columns) in Fig. 5 present seven classifiers trained with the aggregated frames of $\mathrm{CA'}_{\mathrm{T,a9}}$ and evaluated on $\mathrm{CA}_{\mathrm{V,a9}}$, similar to the fifth row in Table 1. All classifiers delivered low accuracies (62–66%) after majority voting, none of them could generalize to the three unknown rerecorded settings in the validation set $\mathrm{CA}_{\mathrm{V,a9}}$. Once the training set comprehended the rerecorded samples of CR_{T} in all four room settings ($\mathrm{CA}_{\mathrm{T,a9}}$), the multi-conditional learning improved the results ($\mathrm{CA}_{\mathrm{af,multi}}$) by 24–30% and achieved 87–95% accuracies (dark blue columns) except for naïve Bayes which had 77.10% after majority voting. As it happened in Figs. 3 and 4, CNN outperformed other algorithms again. The top-5 classifiers were picked for further analysis to select a final model for real-time usage on the robot.

4.4 Classifier selection

Five classifiers remained for the final selection. The support vector machines [33, 35, 37] and KNN [9, 28, 38] have been widely implemented for sound event recognition. The decision tree-based classifiers received less attention [12, 26, 33] while deep learning is the state-of-the-art [10, 16, 27].

Table 2 Classifier properties

	$\mathrm{SVM}_{\mathrm{Lin}}$	DT	RF	KNN	CNN
Accuracy (%)	90.2	89.9	92.6	90.5	95.07
Training time (s)	56	13	213	–[a]	3359
Memory usage (MB)	0.364	64.5	1136.7	63.6	3.5
Prediction time (ms)	0.36	–[b]	–[b]	–[b]	6

The accuracies are $\mathrm{CA}_{\mathrm{mv,multi}}$ results from Fig. 4. The training times are Single-threaded for SVM, DT and RF, but multithreaded for CNN on CPU. Prediction time per feature vector is presented for SVM and CNN because they fit in the memory of the robot

[a] KNN classifier does not have training step

[b] DT, RF and KNN classifiers do not fit in the 64 MB RAM of the robot

To choose the final model, multiple aspects must be considered such as accuracy, training time, memory usage and prediction time. All classifiers in Table 2 show reasonable accuracies between 89.9 and 95.07% what satisfies the first criterion. When the training set size is increased, the DT and RF models grow larger [34, 36]. Although the RF model had the second best accuracy (92.6%) in Table 2, but the memory consumption was over 1 GB after learning 165,872 aggregated frames what was not acceptable for embedded systems. Similarly, the DT model (64 MB) did not fit in the memory. In a previous work, KNN performed closely to SVM in [20] and this classifier does not include a training phase. Nevertheless, the training samples must be cached in the memory what is a clear disadvantage and the bigger the training set, the longer the prediction time of k-nearest neighbor classifier. Namely, KNN with CA_{T} can make one prediction in 23 ms on a high-end AMD FX 8350 desktop CPU which is not enough for real-time processing on the robot and the training set also does not fit in 64 MB RAM. Because of these reasons, DT, RF and KNN were not suitable for eventual tests on the robot.

CNN had the best accuracy (95.07%) after 1 h-long training with moderate memory and CPU usage on the robot (Table 2), therefore, this classifier was selected for onboard deployment. It is worth mentioning that $\mathrm{SVM}_{\mathrm{Lin}}$ is a good alternative to CNN if some accuracy can be sacrificed for negligible memory usage (364 KB) and prediction time (0.36 ms).

The confusion matrix of the CNN model (accuracy: 95.07%, *F*-score: 95.54%, precision: 95.71%, recall: 95.47%) is shown in Table 3 where the cells were left blank if they contained less than 1%. The cicada samples were recognized all the time correctly because of the unique voice characteristics (high pitch) of this animal. Some sound events were challenging for the model because the human laugh is similar to the human speech (6% misclassification), the flatulences are short events which are harder to distinguish from

Table 3 Confusion matrix of CNN model

%	AS	BC	B	CM	C	DB	F	G	L	PA	PI	S	T	V
AS	89						1		1			3		
BC		92		1						4				
B			97								1			
CM		3		94										
C					100									
DB						98								
F	5						86		1			3		
G								99						
L	1								91			6		
PA										97				
PI								3			94	1		
S												98		
T													97	
V														99

The rows show the real events and the columns how they were classified

AS ambient sound, *BC* baby cry, *B* Bell, *CM* cat meow, *C* cicada, *DB* dog bark, *F* flatulence, *G* guitar, *L* laugh, *P* parrot, *PI* piano, *S* speech, *T* traffic, *V* vacuum cleaner

the ambient sounds. Overall, the events were recognized with adequate accuracies (>85%).

The authors executed a preliminary test with the CNN model after multi-conditional learning. This model was deployed to the robot, feature vectors were generated directly from the microphone data. CNN predicted well live input, but the implementation details of final recognition system on a Sony ERS-7 are out of scope in this paper.

4.5 Lossy encoding effects

As it was described in Sect. 2.2, CSIBE-AIBO was recorded with lossy Ogg encoder. This section explains the codec selection procedure and how the VBR settings were determined. The lossy codec influence on CSIBE-RAW is investigated in Fig. 6. A gray line (Ref) shows the baseline performance of the convolutional neural network model with uncompressed data, $\text{MP3}_{\text{train}}$ and $\text{Ogg}_{\text{train}}$ were obtained with half transcoding as well as MP3_{full} and Ogg_{full} with full transcoding. The accuracies of $\text{MP3}_{\text{train}}$ and $\text{Ogg}_{\text{train}}$ matched at 100 kbit VBR otherwise the MP3 encoding caused 0.2–1.2% loss against the Ogg Vorbis results. $\text{Ogg}_{\text{train}}$ lost maximum 0.8% from the reference line even on lower bitrates, therefore, Ogg Vorbis is recommended for half transcoding on any bitrates.

The full transcoding with MP3 caused unforeseen consequences because MP3_{full} was lower by 2.7–7.9% than Ogg_{full}, especially on higher bitrates what contradicts the expectation of good quality over 128 kbit VBR. There might be some special encoding settings in LAME which can make some frequency bands sensitive to the MP3 format. Further investigation is needed later to answer why MP3_{full} had lower accuracy.

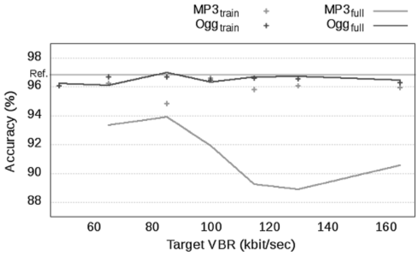


Fig. 6 CNN model performance when both the training ($\text{CR}_{\text{T,a9}}$) and the validation set ($\text{CR}_{\text{V,a9}}$) of CSIBE-RAW were transcoded with lossy codecs (MP3_{full}, Ogg_{full}) and when only the training set was transcoded ($\text{MP3}_{\text{train}}$, $\text{Ogg}_{\text{train}}$), but $\text{CR}_{\text{V,a9}}$ remained in wave format. The results are shown as a function of the target VBR bitrate. The gray reference line (Ref) shows the baseline CNN accuracy with the untouched training ($\text{CR}_{\text{T,a9}}$) and validation sets ($\text{CR}_{\text{v,a9}}$)

Ogg_{full} (blue line in Fig. 6) delivered very similar accuracies compared to $\text{Ogg}_{\text{train}}$ thus the same suggestion applies, Ogg Vorbis codec is advised for full transcoding and the accuracy did not decrease over 128 kbit VBR in comparison with the CNN model built from uncompressed audio.

Although the dataset is limited in this paper, the authors advise Ogg Vorbis to store big audio databases because this format achieved minimal losses in accuracy in both full and half transcoding.

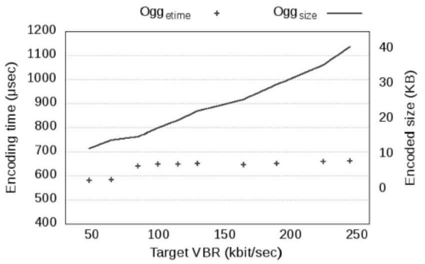


Fig. 7 Processing times on the robot and the encoded sizes of 3-s-long, 16 kHz, stereo audio chunk with Ogg codec. The results are shown as a function of the target VBR bitrate. The crosses are related to the left scale, the lines to the right scale

4.6 Lossy encoding on AIBO

The goals for rerecording samples on AIBO were the small compressed size and the short encoding speed. The small size saved wireless bandwidth and the encoding speed shortened the wait time when the samples were recorded again and collected for multi-conditional learning. Figure 7 shows these variables as a function of different Ogg VBR settings. All processing times (blue crosses) were between 590 and 680 µs, but the produced data chunks were 4 times bigger with 50–250 kbit VBR profiles. To determinate the best compromise between the quality and encoding speed, the Ogg Vorbis codec performance can be compared in Figs. 6 and 7. The higher variable bitrates (>128 kbit) of Ogg Vorbis increase the encoded data size (Ogg_{size} in Fig. 7) without offering additional performance (Ogg_{full} in Fig. 6). Therefore, 128 kbit VBR setting for Ogg Vorbis was the optimal selection to avoid any loss in accuracy caused by audio compression when CSIBE-RAW was rerecorded by the robot for CSIBE-AIBO. Once the audio was encoded with an average compression ratio 1:8, the data were transferred from the robot to a PC in a few 100 ms via wireless network.

5 Conclusion

The paper described how the CSIBE corpora were created for non-overlapping sound event recognition in the robotics field. The samples were mainly gathered from free internet sources to build a redistributable CSIBE-RAW. This database contains 14 sound events where 13 events represented human speech, animal voices, musical instruments and household appliances. One special class modeled the ambient sounds (e.g., knocking, keyboard clicks, paper folding, breathing, steps). CSIBE-RAW was compared to the literature, its size (5954 sample files) was larger than the existing databases for indoor environment and modeling the ambient sounds was also novel.

CSIBE-RAW was rerecorded with a stereo microphone of a robot in four room settings (CSIBE-AIBO) to train acoustic models which were tolerant to reverberant conditions and challenging SNR levels. Multiple experiments were carried out to find the optimal classifier and lossy encoding settings to deploy a real-time capable acoustic model on a Sony ERS-7 robot. The convolutional neural network was the appropriate classifier with multi-conditional learning to reach 95.07% accuracy with unseen audio data from CSIBE-AIBO.

Further contributions of the paper were the reported audio statistics in the recognition system which were robust against the lossy encodings. The lossy Ogg Vorbis and MP3 codecs were studied and the results suggested to select the Ogg Vorbis format with 128 kbit VBR profile.

Future work can evaluate the built MCL model with data recorded from new relative locations and distance to test the generalization power. An other direction can be the introduction of new sound classes to the CSIBE corpora to recognize more environmental events, working out the details of the live sound event recognition on the robot and the investigation of the performance loss with MP3 codec with high VBR profiles. The spectral crest and variance of spectral harmonics were removed because they were altered by lossy encoding. This issue can be analyzed why these statistics are sensitive for MP3 and Ogg Vorbis.

The authors would like to emphasize that the lack of the classifier hyperparameters makes the reported performance measurements hard to interpret because direct comparisons will not be possible with new methods. For example, the NAR dataset evaluation [20] involved SVM classifier, but it is unclear which kernel (linear, radial basis function or polynomial) and hyperparameters were used. The authors of this paper encourage the future works to present the classifier hyperparameters for reproducible research.

Acknowledgements Nokia Foundation provided a Grant (ID: 201510141) for two months in 2015 when the CSIBE-RAW dataset was collected, CSIBE-AIBO was recorded and the initial baseline system was implemented. We want to say special thanks to Toni Heittola and Annamaria Mesaros from Tampere University of Technology for their invaluable comments to overcome some problems.

References

1. Andrew G, Gao J (2007) Scalable training of L1-regularized log-linear models. In: Proceedings of the 24th international conference on Machine learning, pp 33–40

2. Beltrán J, Chávez E, Favela J (2015) Scalable identification of mixed environmental sounds, recorded from heterogeneous sources. J Pattern Recognit Lett 68:153–160
3. Bergstra J, Casagrande N, Erhan D et al (2006) Aggregate features and AdaBoost for music classification. J Mach Learn 65(2):473–484
4. Besacier L, Bergamini C, Vaufreydaz D, Castelli E (2001) The effect of speech and audio compression on speech recognition performance. In: Proceedings of the 4th IEEE international symposium on signal processing, pp 301–306
5. Borsky M, Pollak P, Mizera P (2015) Advanced acoustic modelling techniques in MP3 speech recognition. EURASIP J Audio Speech Music Process 1:1–7
6. Bradski GR, Kaehler A (2008) Learning OpenCV, 1st edn. O'Reilly Media, Newton
7. Bullock J (2007) LibXtract: a lightweight library for audio feature extraction. In: Proceedings of international computer music conference
8. Cakir E, Heittola T, Huttunen H, et al (2016) Polyphonic sound event detection using multi label deep neural networks. In: Proceedings of IEEE international joint conference on neural networks (IJCNN 2016)
9. Chmulik M, Jarina R (2012) Bio-inspired optimization of acoustic features for generic sound recognition. In: Proceedings of 19th international conference on systems, signals and image processing (IWSSIP), pp 629–632
10. Choi I, Kwon K, Hyun Bae S, et al (2016) DNN-based sound event detection with exemplar-based approach for noise reduction. In: Proceedings of detection and classification of acoustic scenes and events workshop (DCASE2016)
11. Chu S, Narayanan S, Kuo CCJ (2009) Environmental sound recognition with time-frequency audio features. IEEE Trans Audio Speech Lang Process 17(6):1142–1158
12. Delgado-Contreras JR, Garcia-Vazquez JP, Brena RF (2014) Classification of environmental audio signals using statistical time and frequency features. In: Proceedings of international conference on electronics, communications and computers (CONIELECOMP), pp 212–216
13. Dennis J (2014) Sound event recognition in unstructured environments using spectrogram image processing. Ph.D. thesis, Nanyang Technological University
14. Foster P, Sigtia S, Krstulovic S, Barkerh J (2015) CHiME-Home: a dataset for sound source recognition in a domestic environment. In: Proceedings of 11th IEEE workshop on applications of signal processing to audio and acoustics (WASPAA)
15. Goldstein EB (2010) Sensation and perception. Wadsworth, p 490
16. Hertel L, Phan H, Mertins A (2016) Comparing time and frequency domain for audio event recognition using deep learning. In: Proceedings of IEEE international joint conference on neural networks (IJCNN 2016). arXiv:1603.05824
17. Hsieh C-J, Chang K-W, Lin C-J (2008) A dual coordinate descent method for large-scale linear SVM. In: Proceedings of 25th international conference on machine learning, pp 408–415
18. Jensen K (1999) Timbre models of musical sounds. Ph.D. dissertation, DIKU report
19. King DE (2009) Dlib-ml: a machine learning toolkit. J Mach Learn Res 10:1755–1758
20. Maxime J, Alameda-Pineda X, Girin L, Horaud R (2014) Sound representation and classification benchmark for domestic robots. In: Proceedings of IEEE international conference on robotics and automation (ICRA)
21. McLoughlin I, Zhang H, Xie Z, Song Y, Xiao W (2015) Robust sound event classification using deep neural networks. IEEE/ACM Trans Audio Speech Lang Process 23(3):540–552
22. Mesaros A, Heittola T, Eronen A, Virtanen T (2010) Acoustic event detection in real life recordings. In: Proceedings of EUSIPCO
23. Ng PS, Sanches I (2004) The influence of audio compression on speech recognition systems. In: Proceedings of 9th conference on speech and computer
24. Ness S, Trail S, Driessen P, Schloss A, Tzanetakis G (2011) Music information robotics: coping strategies for musically challenged robots. In: Proceedings of 12th international society for music information retrieval conference (ISMIR), pp 567–572
25. Nouza J, Cerva P, Silovsky J (2013) Adding controlled amount of noise to improve recognition of compressed and spectrally distorted speech. In: Proceedings of IEEE international conference on acoustics, speech and signal processing, pp 8046–8050
26. Phan H, Maas M, Mazur R, Mertins A (2015) Random regression forests for acoustic event detection and classification. IEEE/ACM Trans Audio Speech Lang Process 23(1):20–31
27. Phan H, Hertel L, Maass M, et al (2016) Robust audio event recognition with 1-max pooling convolutional neural networks. In: Proceedings of 17th annual conference of the interenational speech communication association (INTERSPEECH 2016). arXiv:1604.06338
28. Plinge A, Grzeszick R, Fink G A (2014) A bag-of-features approach to acoustic event detection. In: Proceedings of IEEE international conference on acoustics, speech, and signal processing
29. Pollak P, Behunek M (2011) Accuracy of MP3 speech recognition under real-word conditions: experimental study. In: Proceedings of IEEE signal processing and multimedia applications (SIGMAP), pp 1–6
30. Pollard HF, Jansson EV (1982) A tristimulus method for the specification of musical timbre. J Acust 51:162–171
31. Ruiz-Martinez CA, Akhtar MT, Washizawa Y, Escamilla-Hernandez E (2013) On investigating efficient methodology for environmental sound recognition. In: Proceedings of international symposium on intelligent signal processing and communications systems (ISPACS), pp 210–214
32. Sáenz-Lechón N, Osma-Ruiz V, Godino-Llorente JI (2008) Effects of audio compression in automatic detection of voice pathologies. IEEE Trans Biomed Eng 55(12):2831–2835
33. Salamon J, Jakoby C, Bello J P (2014) A dataset and taxonomy for urban sound research. In: Proceedings 22nd ACM international conference on multimedia, pp 1041–1044
34. Sebbanü M, Nock R, Chauchat J, Rakotomalala R (2000) Impact of learning set quality and size on decision tree performances. Int J Comput Syst Signals 1(1):85–105
35. Stowell D, Stowell D, Benetos E, Lagrange M, Plumbley MD (2015) Detection and classification of acoustic scenes and events. IEEE Trans Multimed 17(10):1733–1746
36. Sug H (2009) An effective sampling method for decision trees considering comprehensibility and accuracy. WSEAS Trans Comput 8(4):631–640
37. Terence NWZ, Dat TH, Dennis J, Siong CE (2013) A robust sound event recognition framework under TV playing conditions. In: Proceedings of signal and information processing association annual summit and conference (APSIPA), pp 1–5
38. Theodorou T, Mporas I, Fakotakis N (2014) Audio feature selection for recognition of non-linguistic vocalization sounds. In: Proceedings of Hellenic conference on artificial intelligence, pp 395–405
39. Tsuruoka Y, Tsujii J, Ananiadou S (2009) Stochastic gradient descent training for L1-regularized log-linear models with cumulative penalty. In: Proceedings of ACL-IJCNLP, pp 477–485
40. Uemura A, Kazumasa I, Katto J (2014) Effects of audio compression on chord recognition. In: Proceedings of international conference on multimedia modeling, pp 345–352

41. Urbano J, Bogdanov D, Herrera P, Gómez E, Serra X (2014) What is the effect of audio quality on the robustness of MFCCs and chroma features? In: Proceedings of 15th ISMIR conference, pp 573–578
42. Wang Y, Neves L, Metze F (2016) Audio-based multimedia event detection using deep recurrent neural networks. In: Proceedings of IEEE international conference on acoustics, speech and signal processing (ICASSP), pp 2742–2746
43. Yamamoto S, Nakadai K, Nakano M, et al (2006) Real-time robot audition system that recognizes simultaneous speech in the real world. In: Proceedings of international conference on intelligent robots and systems (IROS), pp 5333–5338

Paper IV

Kertész, C., (2016). Rigidity-Based Surface Recognition for a Domestic Legged Robot. IEEE Robotics and Automation Letters Journal, 1(1)309-315. doi:10.1109/LRA.2016.2519949

Rigidity-Based Surface Recognition for a Domestic Legged Robot

Csaba Kertész, *Member, IEEE*

***Abstract*—Although the infrared (IR) range and motor force sensors have been rarely applied to the surface recognition of mobile robots, they are fused in this paper with accelerometer and ground contact force sensors to distinguish six indoor surface types. Their sensor values are affected by the crawling gait period, therefore, certain components of the fast Fourier transform over these data are included in the feature vectors as well as remarkable discriminative power is observed for the same scalar statistics of different sensing modalities. The machine learning aspects are analyzed with random forests (RF) because of their stable performance and some inherent, beneficial properties for the model development process. The robustness is evaluated with unseen data after the model accuracy is estimated with cross-validation (CV), and regardless whether a Sony ERS-7 walks barefoot or wears socks, the forests achieve 94% accuracy. This result outperforms the state of the art techniques for indoor surfaces in the literature and the classification execution is real-time on the robot. The above mentioned model development process with RF is documented to create new models for other robots more quickly and efficiently.**

***Index Terms*—Surface Recognition; Random Forests; AIBO**

I. Introduction And Related Work

THE wheels are better options on even surfaces while the legged robots traverse on more difficult terrains. With knowledge about the underlying surface, a legged robot can switch to an efficient gait or adapt its walk speed for optimal locomotion. This paper focuses on the context awareness of domestic robots by predicting six surfaces based with built-in sensors of a Sony ERS-7. Besides the focus on the indoor setting and evaluating the machine learning aspects of past researches, the literature review touches the other conditions in outdoor environments and the use of vision sensors.

The surface recognition is less challenging for outdoor robots because the vibration-based solutions perform better with higher irregularities while the indoor floorings challenge the image classifiers with wider variety of colors and textures. Learning visual cues in a house can enhance a vibration model, but creating a generic texture or color based classifier for all kinds of carpets, tiles and other floorings is an overwhelming task. By these reasons, the terrains were detected with higher accuracies by fused modalities outside compared to the indoor floorings [5] and the vision models suit better in natural environments [4, 17]. These experimental conditions are examined in this paper.

Ojeda and Borenstein [11] experimented with a four-wheeled Pioneer 2-AT on six outdoor terrains and different sensors were explored during the classification process with one training and one testing set. Their neural network produced reasonable performance for the inertial sensors (82.7% accuracy), but the cross-validation was not applicable with their small sample set.

Hoepflinger et al [6] extracted the features from joint motor currents and ground contact force measurements to estimate different terrain shapes and surface properties. A robot leg was fixed to a table in their testbed thus the sensor readings were not affected by robot body oscillations. The model performance of their AdaBoost classifier was not estimated with cross-validation and there is a high chance that these models were overfitted.

By fusing tactile, depth sensors and camera, a six-legged walking robot [15] recognized 12 surfaces with a success rate of 95%. Since only one testing and one training set were evaluated, cross-validation was not performed to estimate the model performance. The feature vector size (174) was much higher than the sample set size (84), therefore, the real accuracy of this method is uncertain with possible overfitting.

Unlike the previous examples, several researches executed appropriate estimations about the model performance with k-fold CV. Hoffman et al [8] explored the terrain discrimination with inertial, tactile, and proprioceptive sensors of a crawling legged robot. Two classifiers (support vector machine (SVM), naïve Bayes) were cross-validated and the performance was estimated at 96.3% with four surfaces (plastic foil, cardboard, Styrofoam and rubber). This result can be compared to the previous work of the author [10] where 93% accuracy was estimated with 10-fold CV when a Sony ERS-7 walked on five surfaces (wood, short carpet, carpet, foam mats, vinyl).

The Sony robots were equipped with noisy, low-end accelerometers (120 Hz) while high-end devices were used in some earlier studies with a sampling rate of 44.1 kHz in [2] and 4 kHz in [5]. A C4.5 decision tree was implemented for AIBO in [13] where the single and pair-joint variances of the three accelerometer dimensions (x, y, z, x-y, y-z, x-z) composed the feature vector and a large sample database was collected. The accuracy of 84.9% was estimated for a model of three surfaces (cement, field, carpet) by 10-fold CV.

OctoRoACH went on three surfaces in [1] and the feature vectors were extracted from inertial measurement unit and force sensor readings. Bermudez et al studied the shortest

Manuscript received: August, 31, 2015; Revised November, 14, 2015; Accepted January, 4, 2016.

This paper was recommended for publication by Francois Chaumette upon evaluation of the Associate Editor and Reviewers' comments. This work is supported by the competence development bonuses of Vincit Oy.

The author is with University of Tampere and Vincit Oy, Tampere, Finland, csaba.kertesz@ieee.org.

Digital Object Identifier (DOI): see top of this page.

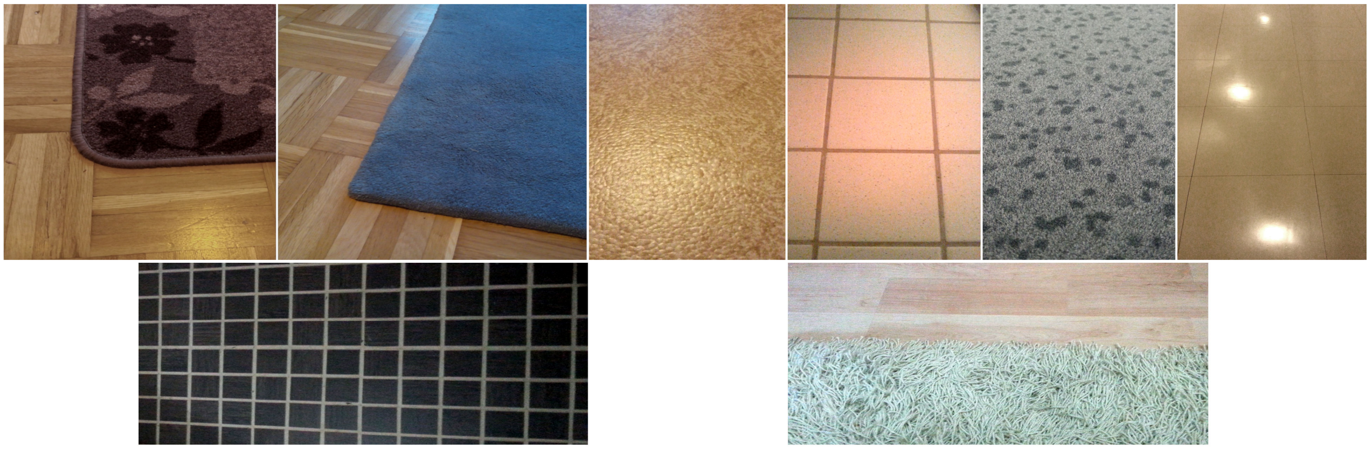


Fig. 1. The images show the following indoor surfaces from top-left: a) Wood flooring and a short carpet, b) Wood flooring and a soft carpet, c) Normal vinyl, d) 8x8cm porcelain tiles, e) Carpeted floor, f) Slippery vinyl, g) 2x2cm porcelain tiles, h) Laminate flooring and a shag carpet.

sampling window (350 msec) to maximize the classifier performance for a running gait. The SVM hyperparameters were optimized by 10-fold CV on a training set and 93.8% accuracy was estimated.

The model accuracy is calculated properly with an evaluation step of unseen data, like in the following works. Degrave et al [3] researched different sensing modalities for supervised and unsupervised classification with five surfaces (blue foil, styrofoam, linoleum, cardboard and rubber). Their experiments found that non-linear sensor fusion and classifier are necessary for good results. The reservoir computing model of tactile and proprioceptive sensors achieved 84.69% accuracy with a validation set.

Tick et al [12] run a wheeled robot on five surfaces (tiled linoleum, ceramic tiles A/B, short carpeted floor and terrazzo). After many statistics were extracted from accelerometer and angular velocity measurements, they studied the sequential forward floating feature selection to build feature vectors for a Linear Bayes classifier. The cross-validation was not executed in this research, however, the built model was tested under practical conditions and the robot achieved 89% accuracy on unseen sensor recordings.

The reviewed literature solved the surface recognition challenges with machine learning methods, but Holmstrom et al applied experience based models to the problem [9]. Specific gaits were evolved by a genetic algorithm for a Sony AIBO robot to walk on plywood board, thin foam, short carpet and shag carpet. After they found some optimal model parameters empirically, tap-delay Adaline neural networks modeled the experience of the leg joints for every gait-surface combination and the robot predicted a surface according to the actual gait. Though a limited sample set was collected, their initial experiments showed promising results and the model accuracy was estimated to 92% with 5-fold cross-validation.

Various locomotion options and sensor combinations have been examined in the past works: wheeled [11, 12, 16, 17], quadruped [3, 4, 8, 9, 10, 13], six-legged robots [15]. Although some studies analyzed a few sensors [9, 11] or the body oscillations [18], this paper does not go into this direction, but the focus is on the classifier model building and the evaluation to provide a better interpretation of the past results from machine learning point of view. Most reviewed studies evaluated their models properly [1, 3, 8, 9, 10, 12, 13], but some did not employed (cross-)validation to get a better estimate of the real accuracy [6, 11, 15]. This paper uses a four-legged Sony ERS-7 robot and the feature vector is composed by the readings of four sensors to train the models. Nine classifiers are examined and random forests are chosen as they have some unique properties for the further analysis. After the features are extracted with fast Fourier transform and statistics, the model accuracy is estimated with 10-fold CV on a training set, and finally, RF is evaluated on a validation set.

II. Experimental Setup

A. Gait and Surfaces

A Sony ERS-7 walked with singular crawling during the experiments to protect the weak servomotors and maintain the stability by having three legs always on the ground [10]. The walk period (φ) was 2400 msec (2.4 Hz) and the speed was around 2cm/sec. The robot traversed on six different, common surface types found in households (Fig. 1):

- 8x8cm porcelain *tiles* and 2x2cm porcelain tiles.
- Lacquer coated *wood flooring* and laminate flooring.
- A bit less rigid normal and slippery *vinyl flooring.*
- *Carpeted floors* with 0.5-1mm thick plastic foam.
- 2-3mm thick *short carpets.*
- 13mm thick *soft carpet* and shag carpet.

Unlike when the robots walked on one example per domestic surface type in [5, 8, 9, 10, 13, 15], AIBO run on multiple examples per type to gather samples in a more generic manner in this paper. The intraclass variability in the dataset were higher in this way and the interclass correlation had a higher chance tough these properties were not studied. Although the body oscillations were mainly influenced by the rigidity and the slipperiness, the surfaces were classified here by the first criterion; for example, soft and shag carpets were placed in the same class. To make the problem even more challenging, the samples were collected with socks and walking barefoot. (The robot worn dog socks with different anti-slip patterns in the experiments, their exact installation can be found in [10].) The idea behind the mixed usage of socks was invented when the author did not find any significant effect on the recognition accuracy with two classifiers trained with no socks/socks cases separately and when the samples were merged together in a single classifier during the initial experimentation. On one hand, the complexity was reduced with collapsed samples, on the other hand, half of the AIBO owners draw anti-slip socks on their robots and the remaining do not use socks at all [10]. It is a clear advantage to use one machine learning model regardless of the owner preference.

B. Sensors

The tactile sensors have been widely used for surface classification [1, 2, 3, 5, 8, 10, 11, 12, 13, 15] to measure the body oscillations during locomotion, and in AIBO, there is a low-cost accelerometer in the torso with a 120 Hz sampling rate.

To the best knowledge of the author, the infrared range sensors were considered by Ojeda and Borenstein [11] for terrain discrimination of mobile robots in the past and they proposed the frequency domain components of these sensors to complement other devices. A built-in IR sensor with a 25 Hz sampling rate on the chest of AIBO was utilized in this paper which was directed to the flooring by 30 degrees.

The advanced force sensors have been often attached to the tip of the robot legs to measure the ground contact forces [3, 6, 8, 15], but the paws in this experiment had a simple two-state contact force sensor with a 10 Hz sampling rate. While proprioceptive sensors (e.g potentiometers) were researched in [1, 3, 6, 8, 9], the ERS-7 model has a force sensor in each leg joint. The previously mentioned papers combined all joints in the sensor readings, however, the author found real discriminative power for the hip joints of the hind legs during the initial experiments. After these hip joints were included in the feature extraction, any other joint had no influence on the classifier performance.

C. Sample Collection

The initial body oscillations (first two walk periods) have an undesired effect on the extracted features as noise, therefore, they were omitted. In [14], the duration of the first step was excluded with similar purpose for a hexapod.

The sampling frequency of the operating system in AIBO is 31.25Hz and the crawl gait is slow, therefore, two walk periods (4.8 seconds) of sensor data were used to extract the feature vectors. This sliding window size is enough to contain a full walk period all the time to catch all body oscillations relevant to the current surface. The author varied the window size in the initial experiments, but the shorter length increased the classifier complexity (random forest size) without any gain in the accuracy. Similar to the this size, Hoffman et al [7] found the 6 seconds-long sensor readings the most accurate with their four-legged robot and an other work [5] concluded the 4 seconds-long window over 1 second. However, shorter windows can suit better for different robots or gaits; Bermudez et al [1] found a 350 msec time window enough for a running hexapod robot to maximize the model accuracy.

30709 samples were collected for the dataset what the author opened under the name of Indoor Surface Recognition Dataset (ISRD - DOI: 10.13140/RG.2.1.3877.5764). The corpus was split into a training (S_T) and a validation set (S_V) randomly in 40%/60% partitions. The training set had 2026 wood flooring, 2109 vinyl, 2214 tiles, 1402 carpeted floor, 2510 short carpet and 2112 soft carpet samples, balancing all classes around 2000 samples. The validation set had 2798 wood flooring, 2809 vinyl, 3128 tiles, 1970 carpeted floor, 2276 short carpet and 5355 soft carpet samples. The first role of S_T was to estimate the classifier accuracy with cross-validation and the second was to build the final models for the evaluation of S_V. Such a large validation set has not been reported in the literature, 75%/25% split was defined in [1] and 84%/16% in [12]. Note that the less fraction of the samples are included for training the more difficult for a classifier to predict the validation set.

III. Feature Vector

Before the classifiers are trained, a feature vector must be defined. This chapter describes how the features were extracted from the sensor data streams.

The feature vectors of surface models contain spectral components and statistical descriptors which are computed from a time window over the raw sensor data. The feature vector size (48) in this paper is on the average compared to previous works:

- Vail and Veloso [13] used 6 features derived from the accelerometer data.
- Bermudez et al [1] suggested 15 statistical features.
- Hoepflinger et al [6] defined 20 features from motor currents and ground contact forces.
- Tick et al [12] selected 68 features out of 864.
- Weiss et al [16] generated 128 FFT components from accelerometer data.
- Walas [15] had 174 features generated from tactile, depth sensors and camera.

The feature extraction is a crucial part of the classification process because the models must comprehend discriminative features to provide good predictions. In the literature, either the researchers selected some scalar statistics without deeper analysis [13, 15] or many features were generated in order to run through feature selection. These practices can lead to the usage of non-relevant features in the first case or ending up different optimal statistics for every sensor in the second case. The author of this paper considered many statistics for each sensor and the best were selected manually after an iterative examination of the feature importances in the initial experimentation. The automated feature selection ended up with various statistics for tactile sensors in [12] however the same were optimal for all sensing modalities in this work.

Albeit the statistical moments have lower computational costs than fast Fourier transform (FFT) analysis and they were preferred in [1], the experiments in this paper did not find these moments sufficient to distinguish the surfaces without the FFT magnitudes. The author believe that Bermudez et al found the moments suitable as their surfaces were very distinct.

Some past works used all components of the Fourier transform [11, 16], some reduced the dimensions with principal component analysis [2] or similar method. Holmstrom et al [9] calculated the FFT on the time series of the proprioceptive sensors in AIBO and the third harmonic peak (~4.5Hz) showed significant difference for multiple surfaces. The F\ourier analysis of tactile sensors in a hexapod robot [15] showed varying magnitudes for more surfaces below 9 Hz and the frequency range 0-4Hz contained most differences, similar to [9]. The author of this paper found that the useful frequency bands were in relation to the walk period, namely, its overtones and the inharmonic partials ($k*\varphi$) hold most information for surface classification where $k \in \{1/16, 1/8, 1/4, 1/2, 1, 3/2, 2\}$. These frequencies were confirmed with feature selection when the first 20 FFT amplitudes of several sensors were added to the feature vector during the initial experiments and the same bands had remarkable feature importances while other bands had negligible. This finding needs a detailed theoretical analysis in the future, but it is not part of the current study.

Every sensor had 150 measurements in two walk periods and the feature vectors were computed over this time window. Six frequency bands contained the proposed overtones and inharmonic partials in the following FFT components: 1st, 2nd, 3rd, 6th, 9th and 12th. Although these bands had good discriminative power for one sensor, but after the same bands were added to the feature vector for new sensing modalities, the gained overall improvements had a decreasing trend. While all six components were worth for the accelerometer (z-axis), the IR sensor had five and the force sensors three. This result may suggest that the FFT analysis of different modalities capture similar discriminative capability caused by the body oscillations from the classifier point of view which implies these oscillations as main influence on the IR sensor readings unlike the surface reflections. This phenomena requires further analysis as well.

A. Accelerometer Sensor

Median, maximum, skewness and root mean square (RMS) amplitude were computed over the sliding window of the accelerometer angles (x, y, z). The robot walk on rigid flooring produces vertical body oscillations, which can be detected in the z dimension [16, 18], while soft surfaces absorb these anomalies. The time series from z axis were transformed to the frequency domain by FFT and the six proposed components were added to the feature vector. 18 features were generated in overall from the accelerometer.

B. Infrared Range Sensor

The IR range sensor on the chest operates within [10; 90 cm] and the robot body oscillations alter these values. The interquartile range (IQR), maximum, skewness, RMS amplitude statistics, the first five proposed and the largest FFT components were added to the feature vector. 10 features were originated from this sensor.

C. Leg Force Sensors

The force sensors in the hip joints of the hind legs were chosen (see Chapter II.B) for feature extraction. The same statistics (IQR, maximum, skewness, RMS amplitude) were calculated again along with the first three proposed and the largest FFT amplitudes. These sensors contributed 16 features.

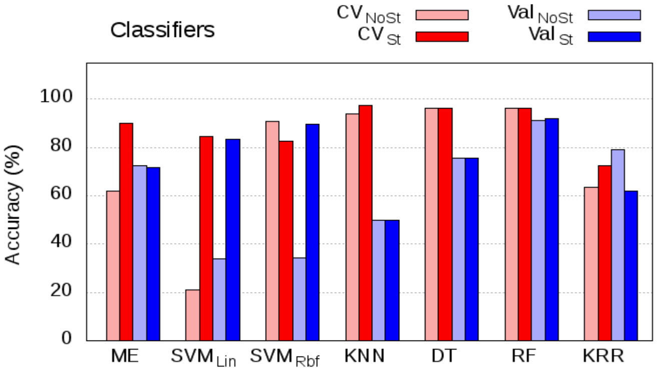


Fig. 2. Every first two bars (*CV_{NoSt}*, *CV_{St}*) show the classifier performances in cross-validation and latter two (*Val_{NoSt}*, *Val_{St}*) for unseen data. Feature standardization was employed for the even bars and there was no feature preprocessing in the odd bars. Classifiers: maximum entropy (ME), support vector machine with linear kernel (SVM_{Lin}), support vector machine with radial basis function (SVM_{Rbf}), k-nearest neighbor (KNN), decision tree (DT), random forest (RF) and kernel ridge regression (KRR).

D. Ground Contact Force Sensors

The four paws of the Sony ERS-7 are two-state buttons. They are pressed more likely when the robot walks on a rigid surface compared to a carpet. Therefore, the pressed durations in a walk period provide a good metric about the rigidity. A simple sum was calculated from these sensors which produced the last four values to the feature vector.

IV. Classifier Selection

According to the well-known no free lunch theorem in machine learning, there is no universal "best" classifier for all problems in the world and different methods can achieve similar, satisfactory results. Nine classifiers were compared in this study (Fig. 2) and the effects of feature standardization were examined during cross-validation and validation phases. (The classifier hyperparameters can be looked up in [10].) The relevance vector machines and naïve Bayes (NB) classifiers achieved low accuracies (< 25%) thus they were omitted in Fig. 2. The latter was unexpected as the Bayes classifiers has been performed remarkable in the literature [8, 10, 12, 16].

Similar to a previous work of the author [10], Weiss et al studied several algorithms with 10-fold cross-validation [16]. Although the decision tree (J4.8 variant) did not yield good results, SVM, KNN and NB had the highest model accuracy estimations in [16]. The author found that NB, SVM and DT were ahead of KNN in [10], but in this paper, SVM, DT, RF, KNN and ME were over 80% in the CV phase.

SVM classifiers are sensitive to the missing feature standardization what is reflected on Fig. 2. SVMs had improved accuracies over 80% in almost every case when the data was standardized. On the other hand, the feature preprocessing caused performance losses in some situations (CV of *SVM_{RBF}*, validation of *KRR*). Although the SVM classifiers had reasonable accuracies in the validation phases, the random forest delivered the most stable performances (91-96%), regardless of applied or absent feature preprocessing. Other classifiers had varying, lower results.

Though neural networks [3, 5, 11, 16], SVMs [1, 8, 10, 15, 16] and decision trees [10, 13, 16] have broad literature in surface classification, the random forests have not been examined at all. As a consequence of the missing experiments, the potentials of this classifier family have not been exploited because the decision trees have limited learning capabilities compared to RF. The random forests were chosen for the further experiments in this paper to investigate the uncovered topics and benefit from the built-in variable importance measures of RF for feature ranking.

A common practice in machine learning to remove the outliers from the database as they can confuse the classifiers, but the author did not find any impact on the accuracy in the initial experiments hence they were left.

Usually it is not expressed in the robotics papers, but the CV results of a classifier can not be matched directly to the model accuracies calculated on unseen data since the first gives only an estimation about the latter. For example, there was CV_{St} >> Val_{St} for ME, DT and KNN (Fig. 2). Therefore, the cross-validation results are compared to the CV values of the literature in Chapter V.A and the model accuracy (Chapter V.B) is presented against the relevant works.

V. MODEL ANALYSIS WITH RANDOM FOREST

The random forests have been implemented for many problems except surface recognition. This paper fills this gap by examining the RF models closely as previous works have not been went into details about the effects of classifier hyperparameters [1, 5, 6, 8, 10, 13, 16], only the feature selections have been researched [3, 11, 12, 15].

The forest dimensions depend on the maximum tree depth and the forest size. $rf_{x,y}$ defines a random forest with maximum tree depth *x* and forest size *y*. The minimum sample count on a leaf for splitting is an other important parameter and it is recommended to set around $|S_T|$ / 100. It was fixed to 100 in these experiments to avoid overfitting.

A. Model Accuracy Estimation

The k-fold cross-validation does not replace the model verification on unseen data, but it gives a reasonable estimate about the model performance. 10-fold cross-validation was run with an $rf_{20,20}$ on the training set to estimate the model performance in this paper and 96.2% accuracy was achieved with six classes. The surface recognition gets challenging by increasing the surfaces as the model must distinguish more and more classes correctly. This novel approach outperformed other methods in the literature because the performance is similar or higher than the estimations of less indoor surfaces and it was better from [5] by 5.7%:

- 3 surfaces: 84.9% in [13].
- 4 surfaces: 92% in [9], 96.3% in [8].
- 5 surfaces: 93% in [10], 96.2% in [3].
- 6 surfaces: 90.5% in [5].

Note that the evaluation above excludes the earlier studies for outdoor terrains [1, 2, 11, 17] since the surface recognition must distinguish more subtle details in the body oscillations on domestic floorings (see in Chapter I) with less surface irregularities. [6, 12, 15] were also omitted by the reason of the missing cross-validation step.

B. Model Accuracy

The cross-validation estimates the model accuracy to some extent, but the results for KNN and DT were far from the real performances in Fig. 2 what warns about the limitations. The models must be built with training samples and tested on unseen data to get a proper measure hence the random forests in this subchapter were constructed with S_T and evaluated with S_V (see in Chapter II.C).

Depending on the random forest parameters, a model can underfit the data if the forest is too small or overfitting happens when too large. Fig. 3 represents how $rf_{i,j}$ models (i,j ϵ [4, …, 80]) were evaluated for their accuracy and memory usage in the function of the maximum forest size and tree depth. The blue-green area on Fig. 3.a shows the underfitting models and the red-orange-green area on Fig. 3.b contains the overfitting models. The area of (i ϵ [10,…, 80]; j ϵ [20,…, 30])

TABLE I. CONFUSION MATRIX OF AN $RF_{20,20}$ MODEL
THE ROWS SHOW THE REAL SURFACES AND THE COLUMNS HOW THEY WERE CLASSIFIED.

%	W	SC	C	V	T	CF
W	**91.57**	0.61	0.54	2.68	4.50	0.11
SC	0.70	**90.99**	3.21	1.49	2.02	1.58
C	0.15	2.22	**90.66**	0.77	2.11	4.09
V	4.13	3.88	0.61	**87.86**	2.85	0.68
T	0.67	0.93	1.28	2.11	**94.66**	0.35
CF	0.36	0.30	1.07	0.86	0.61	**96.80**

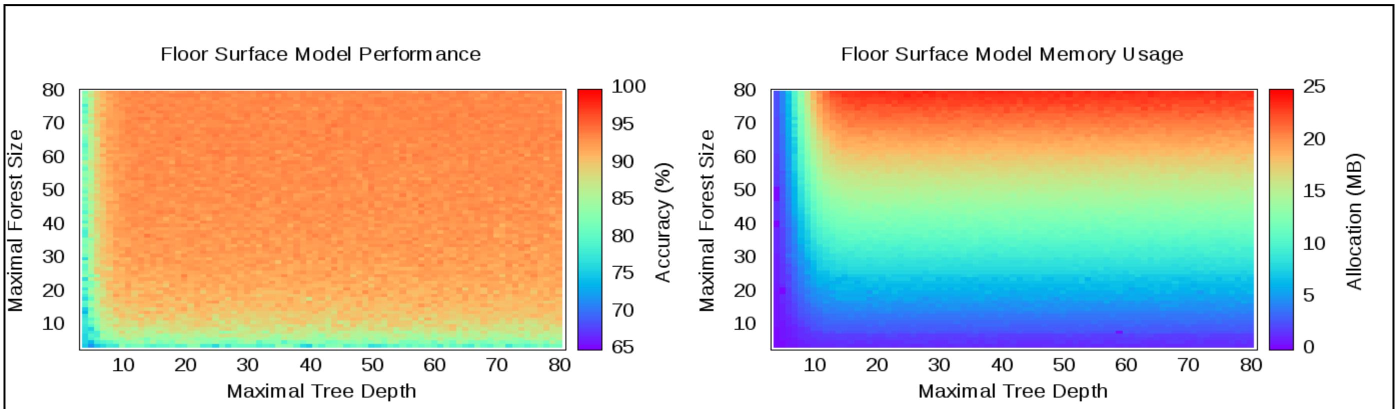


Fig. 3. The diagrams show the random forest model accuracies (3.a) and memory usages (3.b) in the function of the maximum tree depths and forest sizes.

TABLE II. FEATURE IMPORTANCES
THE ODD COLUMNS CONTAIN THE FEATURES, THE EVEN COLUMNS SHOW THE RELATIVE IMPORTANCES FROM H_2O MACHINE SOFTWARE (SCALE: 10^2) IN DECREASING ORDER.
(ACCELEROMETER: A_x, A_y, A_z; INFRARED RANGE SENSOR: IR; FORCE SENSORS OF HIP JOINTS IN HIND LEGS: F_{LH}, F_{RH}; GROUND CONTACT FORCE SENSORS: GF_{LF}, GF_{LH}, GF_{RF}, GF_{RH})

$iqr(F_{LH})$	191	$fft_{max}(F_{RH})$	116	$skew(F_{RH})$	56	$rmsa(A_z)$	27
$rmsa(F_{RH})$	147	$sum(GF_{RH})$	111	$fft_3(F_{LH})$	56	$fft_9(A_z)$	24
$med(A_y)$	147	$max(F_{LH})$	107	skew(IR)	55	$fft_2(A_z)$	24
$fft_1(IR)$	146	$fft_5(IR)$	99	$iqr(F_{RH})$	53	$fft_{12}(A_z)$	23
$sum(GF_{LF})$	146	$max(A_y)$	93	$max(A_x)$	52	$fft_1(A_z)$	22
$rmsa(A_x)$	141	$rmsa(F_{LH})$	75	$fft_3(F_{RH})$	51	iqr(IR)	22
$sum(GF_{LH})$	135	$fft_1(F_{LH})$	74	$skew(A_x)$	48	$fft_3(A_z)$	22
$sum(GF_{RF})$	125	$skew(F_{LH})$	72	$max(A_z)$	33	$fft_6(A_z)$	22
$fft_1(F_{RH})$	124	max(IR)	61	$fft_3(IR)$	33	$fft_4(IR)$	21
$fft_{max}(IR)$	124	$skew(A_y)$	60	$max(F_{RH})$	30	$fft_2(F_{LH})$	20
rmsa(IR)	123	$fft_{max}(F_{LH})$	59	$fft_2(IR)$	30	$fft_2(F_{RH})$	17
$med(A_x)$	123	$rmsa(A_y)$	58	$skew(A_z)$	28	$med(A_z)$	6

TABLE III. ACCURACIES OF VARIOUS SENSING MODALITIES
(ACCELEROMETER: A; GROUND CONTACT FORCE SENSORS: GCF; FORCE SENSORS OF HIP JOINTS IN HIND LEGS: F, INFRARED RANGE SENSOR: IR)

A	GCF	F	IR	A+GCF	A+GCF+F	All
48.1	62.7	69.7	55.2	75.8	87.4	92.0

in Fig. 3.a and Fig. 3.b provides an accuracy plateau of 91-94%. The Table I shows the confusion matrix of an $rf_{20,20}$ model (accuracy: 92.09%, precision: 92.31%) and most misclassifications (orange) happened between classes with similar rigidity. The tiles and carpeted floor had the highest accuracies, therefore, they caused unique body oscillations.

Two earlier works reported model accuracies for five indoor surfaces. Degrave et al [3] achieved 84.69% and Tick et al [12] 89%. The new method of this paper realized a notable 94% for six surfaces and the model generalized over a mixed set of barefoot and sock samples while each class contained several surface examples. This result outperforms [3] and [12] with more surfaces, generalization power and higher accuracy.

C. Feature Importances of Different Sensing Modalities

The variable importances describe the discriminative contributions of the individual features to the model accuracy. A reason for choosing the random forests was the inherent capability to calculate these values after the training phase as Table II shows for an $rf_{20,20}$ model. The author experienced that a feature vector contained unnecessary weak predictors if 5+ features had their relative importances below 10 and removing such variables did not effect the model accuracy. All features in Table II had significant discriminative ability, they differed only in the relative importances to each other.

It was interesting that every modality was significant on average, but the z-axis of the accelerometer produced relative weak discrimination compared to other axes and sensors, its features had low ranks. This result was against the expectation that this axis contains the most descriptive components of the body oscillations [16, 18].

The maximum and 3^{rd} statistical momentum over multiple sensors are in the middle columns of Table II, they provide a stable, average discrimination. The new RMS amplitude statistics (blue) have good performance among the other features and the ground force sensors (yellow) are outstanding despite they are simple two-state sensors. The best FFT features (orange) are principal and maximal coefficients, generated by the infrared range and the motor force sensors. These sensors have been underutilized in the surface recognition, Ojeda et al proposed the infrared range sensors as complementary for inertial sensors [11] what could be originated in more irregularities of the outdoor surfaces. Bermudez et al attached the force sensor to deformable polymeric legs [1] while the force sensors are placed in each leg joints of AIBO.

Table III shows the $rf_{20,20}$ model accuracy when different feature subsets were used for training and validation. The accelerometer (18 features) had the lowest score among the individual sensor features and the leg based features (GCF, F) were strong, similar to [3], while proprioceptive and feet pressure features had higher rankings over the accelerometer in [8]. The accelerometer has been the most popular sensor for surface recognition, but this modality had the lowest relative discriminative power in this paper and in [3, 8]. GCF sensors had good results again, similar to Table II, although they produced only four features. These latter sensors added the biggest contribution to the sensor fusion (A+GCF - 27.7%), force sensors 11.6% (A+GCF+F) and IR 4.1%. The author examined the joint angle sensors in the initial experiments as well, but that modality did not improve the RF models hence they were not included in this work.

D. Computational Requirements

The random forests were coded in C++ with the OpenCV library. A model was built in 1-4 seconds on a first generation Core i7 (1.86Ghz) depending on the forest size. This result outperformed all training times in [16], considering the weaker processor and the larger training set (12373 vs. 9203 samples in [16]) of this paper. The feature extraction with three FFT analyses took 3 msec on a MIPS CPU (576 Mhz) in AIBO and a smaller forest ($rf_{7,5}$) with 90.9% accuracy was selected because of the trade-offs in embedded platforms. This smaller RF predicted a surface in 20-90 ìsec with 833KB RAM.

VI. SURFACE MODEL DESIGN

After the author worked on the machine learning problem of surface recognition and reviewed the available literature, some advices can be given for future researches. These experiences were gathered with a quadruped Sony AIBO, but they may be applicable for other robots:

1. The FFT amplitudes have good discriminative power for sensors of different sensing modalities. The FFT components with the overtones and the inharmonic partials of the walk period are advised (see Chapter III), but after the first sensor is added to the feature vector, only the lower partials of the other sensors contribute improvements to the model performance.

2. Ground force sensors (even simple ones) predict the surface rigidity very well (see Chapter V.C).

3. Recommended statistics for feature extraction: RMS amplitude, IQR, median, skewness and maximum.

4. The applied machine learning is rather a “black art” than exact science nowadays. The author proposes the random forests for the initial experiments and feature selection, but RFs are not the ultimate answer for the surface classification as (legged) robots with different dynamics, sensors or modified feature extraction may need other optimal method. (Usually, powerful features produce similar accuracy with more classifiers.) Although the author selected the features manually, but he believes that the variable importance functions of RF are beneficial to execute this process semi-automatically. The feature vector can be initialized with the FFT amplitudes of a sensor and new features/statistics can be added to the vector with sequential floating forward selection, similar to [12], based on the feature importances. The two main discrete parameters (maximum forest size and tree depth) of RFs are an advantage to control the forest size while many machine learning algorithms have several float hyperparameters. To give an insight to the RF properties, figures can be drawn (Fig. 3) to visualize the sweat spots where these parameters have reasonable accuracy without under- and overfitting. Note that bigger sample sets need longer time to produce these diagrams, up to several hours.

VII. Conclusions

The paper detailed the creation of a random forest model to recognize six indoor surface types. Although the random forests have not been evaluated for surface recognition in the past, they were cross-validated to estimate the model accuracy and the real performance was computed with unseen data in this study. Both results (cross-validation – 96.2%, accuracy - 94%) outperformed the state of the art researches for domestic environments despite the smaller training set, the intraclass variability and the mixed barefoot/socks recordings in the sample sets. The new method had low computational and memory requirements to run the model in real-time on a Sony AIBO. The author found some useful practices what can be applied to the surface model development (see Chapter VI) in the future. Especially, the random forest classifier has some inherent properties how this process can be more effective.

Other contributions were the successful sensor fusion of some underutilized sensors (infrared range, motor force) in the field. A few FFT amplitudes were proposed for surface recognition whose bands were determinated by the overtones and the inharmonic partials of the crawling walk period. The feature selection confirmed the importance of these magnitudes hence future researches with legged robots are encouraged to use these frequencies. It was also found that the classifier captured similar prediction capabilities of the same FFT components of different modalities which rooted on the body oscillations caused by the walk period.

Many statistics were explored to find the maximum, skewness, interquartile range and median beneficial for more sensors. The root mean square amplitude was applied efficiently to all modalities though it has not been considered for surface classification earlier.

Future work can consider more surface types, the examination of traversing the surface edges and the detection of slippery surfaces although there are several limitations in this paper. One gait at a fixed speed was analyzed thus more experiments must be executed with varied conditions since several past studies focused on multiple gaits [8, 9, 12] and speeds [10, 16]. The infrared range sensor was directed to the ground in this paper and the effects of small objects (e.g LEGO bricks) on the floor have not been researched yet.

Acknowledgment

Thanks for the support from my company and the patience of my family when the samples were collected.

References

[1] F. L. G. Bermudez, R. C. Julian, D. W. Haldane, P. Abbeel, and R. S. Fearing, “Performance Analysis and Terrain Classification for a Legged Robot over Rough Terrain,” Int. Conf. on Intelligent Robots and Systems (IROS), pp. 513-519, 2012.

[2] C. A. Brooks and K. Iagnemma, “Vibration-based Terrain Classification for Planetary Exploration Rovers,” IEEE Transactions on Robotics, Vol. 21(6), pp. 1185–1191, 2005.

[3] J. Degrave, R. V. Cauwenbergh, F. Wyffels, T. Waegeman, and B. Schrauwen, “Terrain Classification for a Quadruped Robot,” Int. Conf. on Machine Learning and Applications (ICMLA), 2013.

[4] P. Filitchkin and K. Byl, “Feature-Based Terrain Classification For LittleDog,” IEEE IROS, pp. 1387–1392, 2012.

[5] P. Giguere and G. Dudek, “A Simple Tactile Probe for Surface Identification by Mobile Robots,” IEEE Transactions on Robotics, Vol. 27(3), pp. 534-544, 2011.

[6] M. A. Hoepflinger, C. D. Remy, M. Hutter, L. Spinello, and R. Siegwart, “Haptic Terrain Classification for Legged Robots,” Int. Conf. on Robotics and Automation (ICRA), 2010.

[7] M. Hoffmann, N. Schmidt, R. Pfeifer, A. Engel, and A. Maye, “Using Sensorimotor Contingencies for Terrain Discrimination and Adaptive Walking Behavior in the Quadruped Robot Puppy,” Int. Conf. Simulation of Adaptive Behaviour (SAB), pp. 54-64, 2012.

[8] M. Hoffmann, K. Štěpánová, and M. Reinstein, “The Effect of Motor Action and Different Sensory Modalities on Terrain Classification in Quadruped Robot Running with Multiple Gaits,” J. of Robotics and Autonomous Systems, Vol. 62(12), pp. 1790-1798, 2014.

[9] L. Holmstrom, A. Toland, and G. Lendaris, “Experience Based Surface Discernment by a Quadruped Robot,” Symp. on Computational Intelligence in Image and Signal Processing (IEEE-CIISP 2007), pp. 409-414, 2007.

[10] C. Kertész, “Exploring Surface Detection for a Quadruped Robot in Households,” 14th IEEE Int. Conf. on Autonomous Robot Systems and Competitions (IEEE-ICARSC), 2014.

[11] L. Ojeda, J. Borenstein, G. Witus, and R. Karlsen, “Terrain Characterization and Classification with a Mobile Robot,” J. of Field Robotics, Vol. 23(2), pp. 103–122, 2006.

[12] D. Tick, T. Rahman, C. Busso, and N. Gans, “Indoor Robotic Terrain Classification via Angular Velocity based Hierarchical Classifier Selection,” Int. Conf. on Robotics and Automation (ICRA), 2012.

[13] D. Vail and M. Veloso, “Learning from Accelerometer Data on a Legged Robot,” 5th IFAC/EURON Symposium on Intelligent Autonomous Vehicles, 2004.

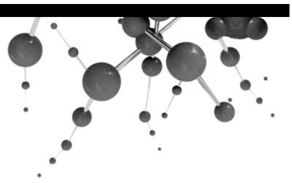
[14] K. Walas, “Tactile Sensing for Ground Classification,” 1st Int. Work. on Perception for Mobile Robots Autonomy (PEMRA), 2012.

[15] K. Walas, “Terrain Classification and Negotiation with a Walking Robot,” J. of Intelligent & Robotic Systems, Vol. 78(3-4), pp. 401-423, 2015.

[16] C. Weiss, N. Fechner, M. Stark, and A. Zell, “Comparison of Different Approaches to Vibration-based Terrain Classification,” 3rd European Conference on Mobile Robots (ECMR), pp. 7-12, 2007.

[17] C. Weiss, H. Tamimi, and A. Zell, “A Combination of Vision- and Vibration-based Terrain Classification,” Int. Conf. on Intelligent Robots and Systems (IROS), pp. 2204-2209, 2008.

[18] Y. Chen, C. Liu, and Q. Chen, “A Vestibular System Model for Robots and Its Application in Environment Perception,” Int. Conf. on Robotics and Automation (ICRA), pp.. 230-235, 2010.



Paper V

Kertész, C., & Turunen, M. (2018). Body State Recognition for a Quadruped Mobile Robot. Proceedings of the IEEE 22nd International Conference on Intelligent Engineering Systems, 323-328.

This is an Accepted Manuscript of an article published in 2018 IEEE 22nd International Conference on Intelligent Engineering Systems (INES), available online: http://dx.doi.org/10.1109/INES.2018.8523877. (©2018 IEEE)

Body State Recognition for a Quadruped Mobile Robot

C. Kertész[*] and M. Turunen[*]

[*] Tampere Unit for Computer-Human Interaction (TAUCHI), University of Tampere, Tampere, Finland

[*] csaba.kertesz@ieee.org, markku.turunen@uta.fi

***Abstract*—The body states must be tracked by the onboard software on the robot to make good decisions. A human can pick up this machine or if the robot encounters anomalies (e.g. fall over) during locomotion, the state changes must be identified to execute the necessary responses. The authors of this paper developed a machine learning model which can recognize four states (normal, pick-up, fall over, poked) of a Sony AIBO robot. A deep neural network classifier with these predictors achieved 98 % accuracy on unseen data and actual test runs on the robot proved the practical use with real-time execution speed. These properties made the proposed method a good candidate for adaption to other legged robots.**

I. Introduction

It is essential to recognize the current robot state to execute the next actions for a mobile robot. To the best knowledge of the authors, the anomaly detection during locomotion is an underresearched topic in the robotics field and the majority of the literature was published in the previous decade, mainly related to the RoboCup competitions.

Hoffmann and Göhring [2] implemented collision detection for legged robots in Robocup. They compared the sensor readings with the actuator commands and their system could detect the collisions with varying success (50-90% accuracy). A limitation was the dependency on the locomotion gait.

Spranger et al [6] developed a biologically inspired approach (Slow Feature Analysis) to recognize postures for a biped robot. Although their main objective was to recognize the basic static postures (e.g. lying, standing) and their state transitions, they could also detect when the robot was fallen. Their work did not include detailed analysis of the recognition accuracy.

Meriçli et al [7] proposed a collision detection framework for omnidirectional motion of a quadruped robot. They examined the temporal accelerometer readings in the frequency domain for regularities (normal motion) and novel situations (collisions). The robot built a probabilistic model while walking without obstruction and used this model to determine any unfamiliar pattern. Their experiments shown quick and successful detection of collisions.

Goswami et al [8] worked on the fall detection for humanoid robots. Their goal was to change the default fall direction of the robot during the fall event to avoid delicate objects or a person. They proposed the fall trigger boundary (FTB) for fall prediction in the robotic systems which is a boundary in the feature space of the robot variables (sensor values, angular momentum). The robot can maintain its balance inside the boundaries though it will definitely fall once the boundary is passed. There was no implementation example for this approach in the simulation environment of [8], but this boundary is learned with machine learning modeling for Sony AIBO in this paper.

Tam and Kottege [9] introduced a robust fall prevention system for bipedal robots. Their Robotis OP2 robot used a walking stick for recovery actions to keep its balance by extending the support polygon on the ground. The Fall Classifier in Robotis OP2 was based on the famous inverted pendelum model to keep the centre of mass inside the support polygon determined by the feet on the ground. The system tracked the inertial measurement unit readings, the angular velocities and the fall angle relative to the walking stick and a fall event happened when the robot control detected unrecoverable instability of the walking sequence. Contrary to [9], the fall detection in this paper was treated like a pattern classification problem in a different approach.

As it can be seen, the literature concentrated on the collision detection, but none of them modeled two situations during human-robot interactions. First, when a human threatens a robot by poking it, and second, when the robot is picked up by person and transported to a new location. The authors of this paper built a machine learning (ML) model to detect these situations and the fall over events for a Sony ERS-7 robot (Fig. 1). The following sections describe how the dataset was created, an optimal ML model was selected and finally tested on the robot.

II. Dataset

A dataset was collected to detect robot state changes. When the robot operates without anomalies and human intervention, it is in *normal state*. The *fall over* events must be detected to recover in these situations and the robot may be under threat if a human starts to *poke* from side. The robot must also recognize if it was *picked up* by a person and it is being transported to a new location. These four states (normal, fall over, poked, picked up) must be recognized by a machine learning method in this paper, therefore, a dataset was collected with a Sony ERS-7 robot. After the feature vectors were generated (see in the next section), 76535 samples were in the dataset: 52612 samples for normal, 12831 for picked up, 1323 for fall over and 9769 for poked state. The normal state has by far the most samples since it involves all activities of a legged robot: walking, sitting, lying, standing, object manipulation, human-robot interaction etc. The picked up and poked sample counts are moderate because they can be easily collected. However, the fall over can stress the robot body regardless of the safe landing on a soft ground thus it cannot be repeated many times for data collection. Furthermore, one fall over generates few samples per event caused by the short duration and these circumstances result a skewed label with 1323 samples.

Fig. 1: Sony ERS-7 robot dogs

The samples were separated randomly to training (S_T) and validation sets (S_V) because the former can be used to run 10-fold cross-validation and train a model while the former is useful to evaluate the built model with unseen data. 57.52% of the samples were part of the training set: 31607 for normal, 5855 for picked up, 694 for fall over and 5870 for poked states. 42.47% of the samples went in S_V: 21005 normal samples, 6976 picked up samples, 629 fall over samples and 3899 poked samples.

III. Feature Vector

The tactile sensors have been widely used for multiple purposes for mobile robots to recognize the directional bias during locomotion [1], collision detection [2] or surface and slope detection [3]. The authors of this paper invented some predictors for body state estimation with a Sony AIBO robot which has a low-cost accelerometer in the torso with a 120 Hz sampling rate. A feature vector representation of the samples is needed for traditional machine learning methods, therefore, these vectors were generated by a 270 msec-long sliding window with 32 accelerometer values per axis and locomotion states. Let us denote the feature vector and the sliding window:

$$FV=\{fv_0,fv_1,\dots,fv_{12}\}, \tag{1}$$

$$A_w=\{a_t,a_{t+1},\dots,a_{t+31}\}, \tag{2}$$

where FV composed of 13 numbers and A_w contains sensor values for one axis (x, y or z). Some statistics were computed for each accelerometer axis separately, but the formulas were otherwise identical. These predictors were defined as:

$$fv_0=iqr\left(A_{w,x}\right),\ fv_1=iqr\left(A_{w,y}\right),\ fv_2=iqr\left(A_{w,z}\right), \tag{3}$$

$$fv_3=min\left(A_{w,x}\right), fv_4=min\left(A_{w,y}\right), fv_5=min\left(A_{w,z}\right), \tag{4}$$

$$fv_6=max\left(A_{w,x}\right), fv_7=max\left(A_{w,y}\right), fv_8=max\left(A_{w,z}\right), \tag{5}$$

where $fv_{0\text{-}2}$ are the interquartile ranges of each axis, $fv_{3\text{-}5}$ are the minimum and $fv_{5\text{-}7}$ are the maximum values. These predictors describe how much the body oscillated inside the sliding window. The rest four predictors contained information about the locomotion state of the robot:

$$fv_9=\begin{cases}1 & : \ W_{forward,t}>0\,msec\\ 0 & : \ otherwise\end{cases},\quad fv_{10}=\begin{cases}1 & : \ W_{backward,t}>0\,msec\\ 0 & : \ otherwise\end{cases}, \tag{6}$$

$$fv_{11}=\begin{cases}1 & : \ W_{turnleft,t}>0\,msec\\ 0 & : \ otherwise\end{cases},\quad fv_{12}=\begin{cases}1 & : \ W_{turnright,t}>0\,msec\\ 0 & : \ otherwise\end{cases}, \tag{7}$$

where fv_9 is non-zero if the robot walks forward, fv_{10} is non-zero if the robot walks backward and $fv_{11\text{-}12}$ are related to the turning directions respectively. $W_{forward,t}$, $W_{backward,t}$, $W_{turnleft,t}$ and $W_{turnright,t}$ are the elapsed time in a certain action and they are beneficial when the locomotion state and the accelerometer statistics are contradicting each other. For example, the robot walks forward and a human picks it up. While $W_{forward,t}$ indicate forward walk in the feature vector, $fv_{0\text{-}8}$ statistics will be out of the ranges of ordinary forward walk.

FV was defined by statistics of accelerometer values and locomotion actions. The following section will describe the classifiers which were trained by FV.

IV. Classifiers

There is no best classifier for all problems and different methods can achieve satisfactory results. After the samples of the training set (S_T) were standardized, six classifiers were evaluated by 10-fold cross-validation (CV) whose results are shown on Fig. 2. The naïve Bayes (NB) and support vector machine (SVM) had the worst results with 73.79% and 76.46%. The SVM classifier used linear kernel with dual coordinate descent method [4] and its hyperparameter was C=0.1. The k-nearest neighbor (KNN) had the best accuracy with 99.60%, followed by the decision tree (DT) classifier with 99.44%. DT had two hyperparameters, the max depth was set to 20 and the tree node sample limit was 1. The random forest (RF) had lower result (98.09%) than DT which can be caused by the task complexity. A random forest is a collection of decision trees and RF can learn a problem better than one DT if the learning capacity in one DT is not enough for the complexity.

The deep learning is a major research area in machine learning nowadays and breakthrough results appear with this approach in many robotics fields. The deep neural network (DNN) in this study has two hidden layers with 20-20 neurons and leaky ReLu activation functions. The output layer has 4 neurons with softmax function to calculate the classification result. The adagrad adaptive gradient algorithm was run on the network with batch size 64 for 50 epochs. Although somebody can argue that this network topology is not bigger than the classical neural networks from the nineties, but the recent developments in the learning algorithms and other parts made the new networks outperform the old in even simple tasks. The authors of this paper found small neural networks surpassing traditional methods with handcrafted features lately in [5]. DNN had a moderate result (88.95%) in the 10-fold cross-validation.

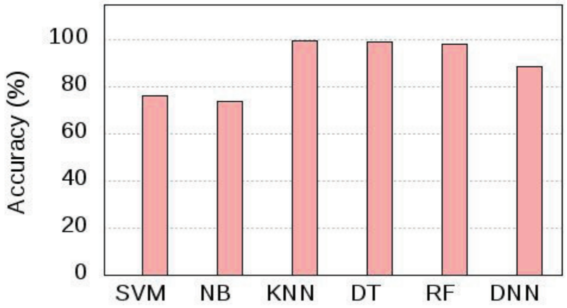


Fig. 2: 10-fold cross-validation results of six classifiers: support vector machine (SVM), naïve Bayes (NB), k-nearest neighbor (KNN), decision tree (DT), random forest (RF) and deep neural network (DNN).

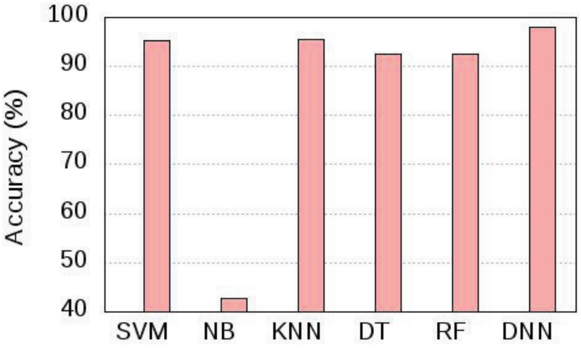


Fig. 3: Model evaluation results with six classifiers: support vector machine (SVM), naïve Bayes (NB), k-nearest neighbor (KNN), decision tree (DT), random forest (RF) and deep neural network (DNN).

Table I. Confusion matrix of deep neural network model results. The rows show the real body states and the columns how they were classified.

	Normal	Picked up	Fall over	Poked
Normal	99.28	0.18	🟨	0.53
Picked up	1.90	97.40	🟨	0.68
Fall over	13.35	🟨	86.64	🟨
Poked	3.61	2.30	🟨	94.07

Table II. Memory and runtime requirements of five models.

	SVM	KNN	DT	RF	DNN
Model accuracy	95.35%	95.42%	92.59%	92.59%	98.01%
Memory consumption	2 KB	2.5 MB	117 KB	181 KB	28 KB
Execution time	2 usec	1.6 msec	3 usec	3 usec	20 usec

After the cross-validation was run on the training set (S_T), models were built with the whole S_T set and their performance were evaluated with unseen data (S_V) in the next section.

V. Model Selection

Models were built upon the training set (S_T) and the accuracies were calculated by the samples of the validation set (S_V). Fig. 3 shows the results which did not follow the trends of the cross-validation in Fig. 2. This phenomenon is originated in the fact that the cross-validation gives an estimation about the real classifier performance on unseen data and the reality can be far from the CV accuracies. As it can be seen, SVM had the second worst CV result, but it achieved third best accuracy with 95.35%. NB had the worst result in Fig 2. and this classifier collapsed in the model evaluation with 42.65% accuracy. The k-nearest neighbor had the best CV value, but it reached the second place in Fig. 3 with 95.42%. DT and RF had good cross-validation results and they achieved performance over 90% with 92.59% in Fig. 3. The deep neural network was not the best in the cross-validation round, nevertheless, it was the top-performer in the model evaluation with 98.01% accuracy. This result suggests, similar to [5], that the new generation of neural networks can outperform the old methods by margin in even small problems. While the traditional algorithms have still the advantage of the quicker learning speed and a possible smaller memory footprint, it is a disadvantage that they need hyperparameter optimization even for small problems to achieve the best performance. On the other hand, the authors experienced that deep neural networks can be trained with handcrafted features, default hyperparameters and simple multilayer perceptron topology to learn a training set with good generalization power out of the box.

The confusion matrix of the DNN model is shown in Table I. As expected, the normal samples had the best accuracy (99.28%) since this state had the most samples in the dataset. The picked up and poked labels were recognized well with 97.40% and 94.07% accuracies although 1-3% misclassifications happened in a few cases. The fall over had the least, 86.64% accuracy and 13.35% of these samples were misclassified to normal state. Since none of the other labels were misclassified as fall over and fall over samples were misclassified *only* as normal state (yellow and gray cells in Table I), the authors believe that the predictors were strong to identify the fall over situation. But when a fall over event starts, the samples are similar to the normal status before the robot turns over its center of gravity and falls with accelerating speed. Since the sliding window is 270 msec long to compute a feature vector, some time is necessary until the fall over event dominate the accelerometer statistics inside the sliding windows. This is a possible explanation for this misclassification.

The memory requirements and the execution times of all classifiers are listed in Table II except naïve Bayes which failed to build a satisfactory model. The k-nearest neighbor has the highest memory consumption and prediction time because this method stores all training samples in the memory and use them in each prediction step. The decision tree and random forest had moderate memory usage, their execution times were quick, but their accuracies were relative low compared to the other classifiers. The SVM achieved the best memory and execution time results, therefore, this classifier is recommended for robots with microcontrollers where every KB and microsecond count. However, the deep neural network had small resource needs paired with the best model accuracy thus this classifier is the best choice to run in microsecond-scale on any embedded CPU.

VI. Evaluation on The Robot

The deep neural network outperformed the other classifiers with unseen data from the validation set (S_V). However, this evaluation does not depict exactly how the model will work under real conditions. To address this problem, the DNN model was deployed to a Sony ERS-7 to inspect how the classifier behaves in some typical situations without additional filtering.

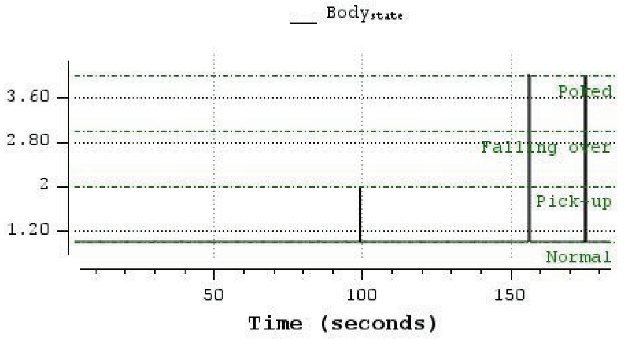


Fig. 4: A normal runtime with basic postures and locomotion. External events did not cause any disturbances for the robot.

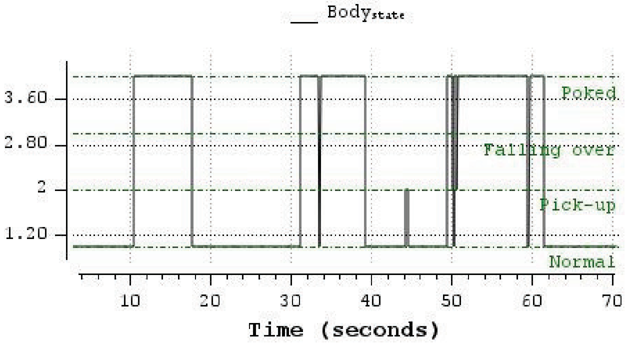


Fig. 5: The robot was poked in sitting, standing and lying postures.

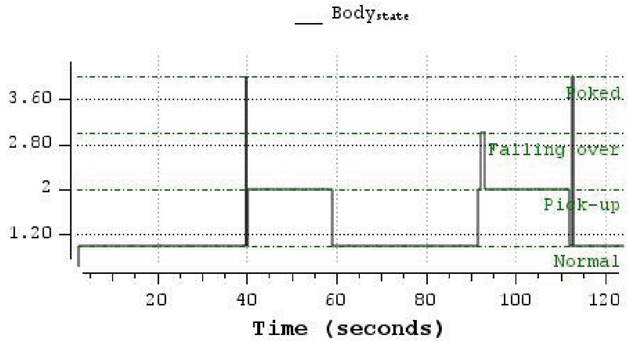


Fig. 6: The robot was picked up from lying posture and

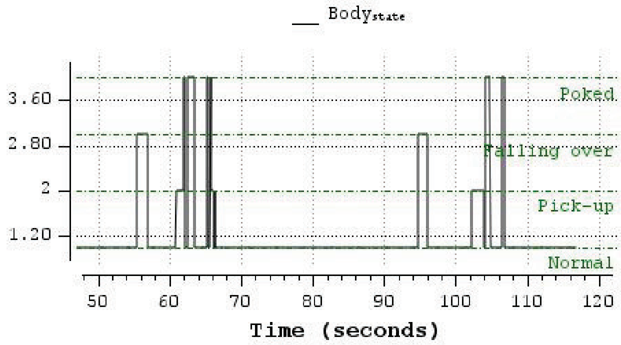


Fig. 7: An example runtime when the robot falls.

In the first case, the robot completed usual activities without any disturbances. After transitioning between lying, sitting and standing postures, the robot walked forward, backward and turned in both directions. The recognized body states are shown on Fig. 4. As it can be seen, the classifier detected the normal operation correctly almost all the time. There were three very short periods when picked up and poked events were identified for a moment.

The robot can be disturbed by poking in static postures (sitting, lying, standing). Fig. 5 shows that the robot was poked between 10-18 seconds in lying posture. After the transition was finished to sitting posture, the robot was poked between 31-39 seconds. And after standing up, the robot was tossed again between 50-62 seconds. As we can see, the long poking events were recognized correctly and only a very short picked up event was misclassified around 45 seconds during the transition between sitting and standing postures.

Fig. 6 demonstrates picked up events. The robot was in lying posture when a person picked it up and walked around between 40-60 seconds. The robot was placed back to the floor later before it stood up and started to walk forward. The person picked it up again and carried between 91-113 seconds. The picked up states were recognized correctly except two short poked and one fall over events.

Fig. 7 shows two fall over events during locomotion. The robot fallen to the left side while it walked forward after 55 seconds and to the right side after 95 seconds. Till the robot was turned back to lying posture, the misclassified poked and picked events happened for longer periods compared to the previous situations. These events can be easily suppressed until the lying posture is reached after falling down.

As Fig. 4-7 showed the deep neural network model was robust and misclassifications did not happen for significant periods. These results suggest that the model can recognize all states correctly under realistic conditions after a minimal filtering.

VII. Conclusion

This paper described a machine learning approach to detect sudden changes in the robot state such as fall over, poking and being picked up by a human. A legged robot was used for data collection and several classifiers were challenged to find the best performing for embedded use. The deep neural network achieved 98% accuracy and its low resource usage made suitable for running in embedded robotics systems. Although the built model processed each data frame separately without temporal fusion, actual tests on the robot shown that the proposed method was applicable under real circumstances with minimal false positives.

Future work can include the further analysis of temporal machine learning approaches (e.g. recurrent neural networks) with raw sensor data since the current solution operates with handcrafted features. Another direction can be to add new states for the model or study the applicability of the predictors with other mobile robots.

References

[1] D. J. Huang and W. C. Teng, “A Gait Based Approach to Detect Directional Bias of Four-Legged Robots’ Direct Walking Utilizing Acceleration Sensors,” *Intl. Conf. on Knowledge-Based and Intelligent Information and Engineering Systems*, 2007, pp. 681-688.

[2] J. Hoffmann and D. Göhring, “Sensor-actuator-comparison as a basis for collision detection for a quadruped robot,” in *Robot Soccer World Cup*, 2005, pp. 150-159.

[3] Y. Chen, C. Liu, and Q. Chen, “A vestibular system model for robots and its application in environment perception,” *Intl. Conf on Computing, Control and Industrial Engineering (CCIE)*, 2010, pp. 230-235.

[4] C. J. Hsieh, K. W. Chang, C. J. Lin, S. S. Keerthi, and S. Sundararajan, “A dual coordinate descent method for large-scale linear SVM,” *25th Intl. Conf. on Machine Learning*, 2008, pp. 408-415.

[5] C. Kertesz and M. Turunen, “Common Sounds in Bedrooms (CSIBE) Corpora for Sound Event Recognition of Domestic Robots,”, in *Intelligent Service Robotics*, forthcoming/in press, 2018.

[6] M. Spranger, S. Höfer, and M. Hild, “Biologically inspired posture recognition and posture change detection for humanoid robots,” *IEEE Intl. Conf on Robotics and Biomimetics (ROBIO)*, 2010, pp. 562-567.

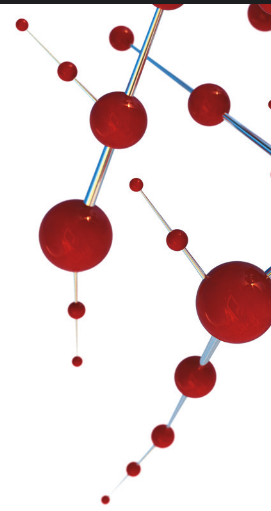
[7] T. Meriçli, C. Meriçli, and H. L. Akin, “A robust statistical collision detection framework for quadruped robots,” in *Robot Soccer World Cup*, 2008, pp. 145-156.

[8] A. Goswami, S. K. Yun, U. Nagarajan, S. H. Lee, K. Yin, S. Kalyanakrishnan, “Direction-changing fall control of humanoid robots: theory and experiments,” in *Autonomous Robots*, 36(3), 2014, pp. 199-223.

[9] B. Tam, N. Kottege, “Fall avoidance and recovery for bipedal robots using walking sticks,” in *Australasian Conf. Robotics and Automation*, 2016.

The main aim of this doctoral thesis was to investigate on how to involve a community for collaborative artificial intelligence (AI) development of a social robot. The work was initiated by the author's personal interest in developing the Sony AIBO robots that have been unavailable on the retail markets, however, user communities with special interests in these robots remained on the internet. An active online community of Sony AIBO owners was approached to investigate factors to engage its members in the creative processes.

There are significant contributions in this dissertation to robotics. First, the long-term robot usage was not studied on a years-long scale before and the most extended human-robot interactions analyzed test subjects for only a few months. A questionnaire investigated the robot owners with 1-10+ years-long ownership in this work and their attitude towards robot acceptance. The survey results helped to understand the viable strategies to engage users for a long time. Second, innovative ways were explored to involve online communities in robotics development. The past approaches introduced the community ideas and opinions into product design and innovation iterations. The community in this dissertation tested the developed AI engine, provided inputs for further development directions, created content for the actual AI and gave their feedback about product quality. These contributions advance the social robotics field.



Dissertations in Interactive Technology
Number 36

ISBN 978-952-03-2442-1
ISSN 1795-9489